

SQ Sequence 964 AA:

Query Match 98.2%: Score 4994.5; DB 21; Length 964;
 Best Local Similarity 98.4%: Pred. No. 0;
 Matches 950; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

QY 1 MKKAFEPFLIGSLGSLAEVPSRIETLMPNSVDPPTKESLNKISLTGDTNLTNCYLON 60
 DB 1 mkkafefllignslsgrlarevpsriltmpnsvdpptkeslnksistgdtmnlncylch 60
 QY 61 LRYIIAIILOKTPNEGAAVYITDYLSEFFDQEGYIFAKNLTPESGGAIGYASNSPTVEI 120
 DB 61 lryiiaiioktpnegaaavlitdylsefdtqegiyifaknltpesggaigyaspnsptvel 120
 QY 121 RDTICGVIFENNTCCRPFFSSPNNAVKKIRGGALHAONLTIINHHDDYVGFKNFSYR 180
 DB 121 rdticgvifennntccrpfsspnnaavkkirggalhaonltiinhhddyvgfknfsyr 180
 QY 121 rdtlsgpvllennccrclftwrnpysaa-dkireggaihaqnlyinhhdvvgfknfsyvg 179
 QY 181 GGAISTANFEVVSSENSCFLFMDNICIOTNTAGKGAIVAGTSNFSNCDLFFINNC 240
 DB 180 ggaistantfevvsenscflfmdniciotntagkgaiyagtsnfsnncdlffinnnc 239
 QY 241 CAGGAIFSPICSLTGNRGNIVEFYNNRCFKNVETASSEASDGAIKYTRLDVYTGNGRIF 300
 DB 240 caggaifspicsltgnrgrnivefynnrcfknvetasseasdgaikytrldvyngrif 299
 QY 301 FSDNTITKNGGAIYAPVLTVDNGPTYFINNIANNKGAIVYDGTNSKISADRAHAIIRN 360
 DB 300 fsdntitknggaiyapvltvdngptyfinniannkgaivydgtnskisadraliirf 359
 QY 361 ENIVNTVNTANGSTISANPRRNATIVASSGEILLGAGSSQMLIFDYPLEVSNAGVVS 420
 DB 360 enivntvntangstisamprrnativassgeilllgagssqmlifdyplevsnagvs 419
 QY 421 FKKEADQTSVVFSGATVVSADFHORNLOTKTPAPLTLSNGFLCIDHAQLTVNRTQTG 480
 DB 420 fkkeadqtsvvfsgatvvsadfhornloktktpapltlsngflcidhaqltvnrtqtg 479
 QY 481 GVVSLGNGAVLSCYKNGACNSASMSSTIKHIGLNLSTLKSQAEIRLLWEPYNNNNY 540
 DB 480 gvvsllngavllscykngtgdssasaeltkhiglnlsslksqaeirllwepynnnny 539
 QY 541 TADTATFSLSPVKLSLIDYGNSPYESTDLTHALSSQMLISSESDMOLRSDDDDFSG 600
 DB 540 tadtaetfslsvklsliddygnspyescdlthalsqpmllisseadmqldsemdtsg 599
 QY 601 LNVPHYMGGLTWGMAKTQDEPPASSATITDPKANKRFHRTLLTLWLPAGVYPSFKHRS 660
 DB 600 lnvphymggltwgmaktqdeppassatitdpkankrfhrtlltlwlpagvypsfkhrs 659
 QY 661 PLIANTLMGNMLATBSLKNASALTPSDHPFMCITGGGCMVYQDPRENNHGFHRRSSG 720
 DB 660 pliantlmgnmlatbslknasaltpsdhpfcitgggcmvyqdprennhgfhrrssg 719
 QY 721 YSAGMIAGOTHFSLKFSQTYTKLNERVAKNNVSSKNYSCOGEMLSLOEGFLTLYVG 780
 DB 720 ysagmiagothfslkfsqtytklnervaknnvssknyscogemlsloegflltlyvg 779
 QY 781 YSYGDHNCHEFTYTOGENLTSGTFRSQTMGAVFEDLPKKPEGSTHILITAPLGAIGIYS 840
 DB 780 ysygdhnccheytogenltsgtfrsqtmgavfედlpkkpegsthillitapligaiyis 839
 QY 841 SLSHFTEVATYRSTSTKPLINLVPIGVKGSFMNATORPQAWYELAYQVPIYAOBEG 900
 DB 840 slshftevatyrststkpplnlvlpigvkgssfmnatorpqawyelayqvpiyaopeg 899
 QY 901 IATQLLASGIVFGSGSPSRHMASYKISQOTOPLSMLTLHFQYHGFYSSTSCNTLNGE 960
 DB 900 iatqllasgivfsgsgpsrhamasyklsqotoplsmltlhfqyhgfysstscntlnge 959
 QY 961 IALRF 965
 IIILII

DB 960 ialrf 964

RESULT 3
 BI3640
 ID BI3640 standard; Protein: 977 AA.
 XX
 AC BI3640:
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE C. trachomatis pmpe gene amino terminus minus signal sequence protein.
 XX
 KW Chlamydial infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 OS Chlamydia trachomatis.
 XX
 PN MO200034483-A2.
 XX
 PD 15-JUN-2000.
 XX
 PE 08-DEC-1999; 99MO-US29012.
 PF
 PR 08-DEC-1998; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skelky YAW, Fling SP, Jen S, Stromberg EJ;
 XX
 DR WPI: 2000-431303/37.
 XX
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence .
 XX
 PS Claim 2: Pages 210-212; 256pp; English.
 XX

The present invention relates to new nucleic acid sequences and the proteins encoded by the nucleic acid sequences. The encoded proteins comprise an immunogenic portion of a Chlamydia antigen. The encoded proteins are useful for the serodiagnosis and treatment of Chlamydia infection. Chlamydiae are intracellular bacterial pathogens that are responsible for a wide variety of human infections. C. trachomatis infection is one of the most common sexually transmitted diseases and can lead to pelvic inflammatory disease (PID), resulting in tubal obstruction and infertility. Trachoma due to ocular infection with C. trachomatis is the leading cause of preventable blindness worldwide. C. pneumoniae is a major cause of acute respiratory tract infections in humans and is also thought to play a role in the pathogenesis of atherosclerosis and coronary heart disease. The present sequence is a protein isolated in the present invention.

SQ Sequence 977 AA:

Query Match 96.7%: Score 4919.5; DB 21; Length 977;
 Best Local Similarity 98.4%: Pred. No. 0;
 Matches 935; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

QY 16 GLAREVPSRIETLMPNSVDPPTKESLNKISLTGDTNLTNCYLONRYIIAIILOKTPNEG 75
 DB 29 glarevpsriltmpnsvdpptkeslnksistgdtmnlncylonlryiiaiioktpneg 88
 QY 76 AAVTTIDVLSFFDQEGYIFAKNLTPESGGAIGVAPNSPVEIRDTIGVPIFENNTCC 135
 DB 89 aavttldvlsfdtqegiyifaknltpesggaigyaspnsptvelrdtlgrpvlfenncc 148

```

Db      269 regvlfnnmgsgnggallhaksillikengpyvflnntatrgallnlisagsgngsfilisa 328
OY      353 DRHAIFENENITVNTNANGSTSANPPRRNAITVASSSGEILLGAGSSONLIFEDPIEV 412
        |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      329 dngdillfnn-----taskhalnpyrna-h-stpmnqlqigarpgyrlyfbydpleh 379
OY      413 S-NAGVSVSFNKEADOTGVSVFSGATVNSADFHORNLOTKTPAPRLTLNSGFLICEDHAOL 471
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      380 elpsfpllnfletghtgvltsgevhqnfcdemfflylntselrgvlayedgagl 439
OY      472 TVNRFOTGCVVSLGNGAVLSCYKNGAGN-----SASNSITLKHIGLWLSSTLKS 523
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      440 acykffqgcllllqggavilt-----agtlpssctptvgstltlnhaidpilsifq 495
OY      524 AEPLLWPTNNSNNYXTADTAFTSLSDVKLSLIDDYNSPFEESTDLTHALSSQMLSI 583
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      496 agapkiwlypctgtctytedsnptltisgt-lltnsmedpydsldshslekpyllyl 554
OY      584 SEASDNQLRSDMDPFSGLNV-PHYGNOGLMTWGMWAKTODPEPASSATITDPK---ANRF 639
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      555 vdaaagklnsqldstlnsgehygyglwstcywet-----tltnpssllganlk 606
OY      640 HRTLTLTWLPACIVSPKRRSPULANTLWGNMLLATESLKNSAELTPSDHPWGITGC-- 697
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      607 hlllyanwsplygrpherrgefltnalwgsaytalaglshsiss-----wdeekgna 658
OY      698 ----GLGMWVYDDPRENHGPFHMRSSGVSAGMIA--GOTHTSLKFSOTYTKLNEVYAKN 751
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      659 aslqglqlllvhbkdkngflgfrtslmngysatleatsgspnslglaqlfsakakehsqn 718
OY      752 NVSSKNY---SCOGEMLFSLQECFLTLTKLVGL-YSYGDHNCNHFYQCGENLISQCFRQ 807
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      719 stsshyfsgmeclak--yslqrvrls--vsalymtsehtltmqgllegnsgsfnnh 774
OY      808 TNGGAVFELPMKPRGSTIILTAPELGAIGYSSLSHTEVGAVPPRSSTKTPRLINVLVP 867
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      775 llaagascvflpqhges-ldlypfltalalrqnlaafesgdharefahlprrltdvslp 833
OY      868 IGVKGSFNMATORPOAWTVELAYOPVLYRQEPDIAQLASGIMFSGSSPSRRHMSYK 927
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      834 vltiraekwhhvpvlwlelsyrslllyqdpelnsklillsqclwtlqatpvtynalglk 893
OY      928 ISOQTOLPSWLTLLHFOYHGVSSTFCNLYLNGEIALRF 965
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      894 vktmqvfpkvlslsdsadissstlsbhylnvascmrf 931
RESULT 5
Y35082
ID      Y35082 standard; Protein: 940 AA.
XX
AC      Y35082;
XX
DT      13-SEP-1999 (first entry)
XX
DE      Chlamydia pneumoniae surface exposed polypeptide.
XX
XX      Respiratory disease: pneumonia; bronchitis; heart disease: sarcoidosis;
KM      sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW      vaccine; neutralising epitope.
XX
OS      Chlamydia pneumoniae.
XX
PN      MO9927105-A2.
PD      03-JUN-1999.
PF      20-NOV-1998; 98MO-1B01890.
PR      04-NOV-1998; 98US-0107078.
PR      21-NOV-1997; 97FR-0014673.
XX
PA      (GEST ) GENSET.

```

```

XX      Grifftals R.
PI
XX      WPI: 1999-357842/30.
DR
XX      Genome sequence of Chlamydia pneumoniae
PT
XX      Page 965-967; Disclosure: 1912pp; English.
XX
CC      Y34584-Y35879 represent the proteins encoded by all the open reading
CC      frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC      pneumoniae causes respiratory disease such as pneumonia and
CC      bronchitis and is thought to be a contributing factor in heart
CC      disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC      nodosum or pharyngitis. The polypeptides encoded by the open reading
CC      frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
CC      immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC      nucleotide sequences can also be used as immunogenic compositions,
CC      especially where the vector directs the expression of a neutralising
CC      epitope of C. pneumoniae.
XX
SQ      Sequence 940 AA:
XX
Query Match 22.0%; Score 1120.5; DB 20; Length 940;
Best Local Similarity 30.9%; Pred No. 1e-74; Indels 109; Gaps 31.
Matches 308; Conservative 163; Mismatches 418;
OY      6 FFFLIGNSLSLAREVPSRIFLMPNSVDPPTRESLSNKLSTLGTDRHNL-----NC 56
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      14 flfvlan--egjq|pletyltispeygaapq-----vgfthnqgdlayvghnd 61
OY      57 YLDNLRIYLAITLQKTPNECAVTTIDYLSFPTQKEGIFAKNLPREGGAIGVSPNSP 116
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      62 lldykky-----rsmgaltcknll--lsenlgnvlfekhvcpnsggal-yaagnc- 110
OY      117 TVEIRDTICPVLFENNCCRPFTSSNPNAVNKIREGCAIHNQNLINNHHDVVGPMKMF 176
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      111 -----tlsk--ngyafatnlvsdnpelatagsll-9gaalfaincsltnlbggtfvdl 161
OY      177 SYVRGCAISTANTFVVSENOQCFLEMDNICIOTNTAGKGAIVAGTSNSESNCDELPEI 236
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      162 alnkgalyletnlsikdnkxprlllkqtral--nedslygglysgnslnlegnsaqitl 219
OY      237 NNACCAGAIIFS-PICSLTGNRGNIVFYNNRCFKVETIASSEASDGAIKVTRIDVTGN 295
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      220 snsssgggllstqltcltissnkkllsensaflan--nygsnfnpgggllttfcilnn 277
OY      296 RGRIFFSODITKNYGAIAPVVTLDVNGPPTYFNNIANKKGAIYID-CTNSK--ISA 352
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      278 regvlfnnmgsgnggallhaksillikengpyvflnntatrgallnlisagsgngsfilisa 337
OY      353 DRHAIFENENITVNTNANGSTSANPPRRNAITVASSSGEILLGAGSSONLIFEDPIEV 412
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      338 dngdillfnn-----taskhalnpyrna-h-stpmnqlqigarpgyrlyfbydpleh 388
OY      413 S-NAGVSVSFNKEADOTGVSVFSGATVNSADFHORNLOTKTPAPRLTLNSGFLICEDHAOL 471
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      389 elpsfpllnfletghtgvltsgevhqnfcdemfflylntselrgvlayedgagl 448
OY      472 TVNRFOTGCVVSLGNGAVLSCYKNGAGN-----SASNSITLKHIGLWLSSTLKS 523
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      449 acykffqgcllllqggavilt-----agtlpssctptvgstltlnhaidpilsifq 504
OY      524 AEPLLWPTNNSNNYXTADTAFTSLSDVKLSLIDDYNSPFEESTDLTHALSSQMLSI 583
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      505 agapkiwlypctgtctytedsnptltisgt-lltnsmedpydsldshslekpyllyl 563
OY      584 SEASDNQLRSDMDPFSGLNV-PHYGNOGLMTWGMWAKTODPEPASSATITDPK---ANRF 639
        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      564 vdaaagklnsqldstlnsgehygyglwstcywet-----tltnpssllganlk 615
OY      640 HRTLTLTWLPACIVSPKRRSPULANTLWGNMLLATESLKNSAELTPSDHPWGITGC-- 697

```

RESULT 7
ID Y92832 standard; Protein: 1000 AA.
XX Y92832:
AC
XX 07-SEP-2000 (first entry)
DT
XX C. pneumoniae CPN100626 full-length antigen.
DE
XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW anti-arteriosclerotic; vaccine.
XX Chlamydia pneumoniae.
OS
XX MO200024765-A2.
PN
XX 04-MAY-2000.
PD
XX
XX 28-OCT-1999; 99WO-CA000992.
PF
XX 28-OCT-1998; 98US-0106034.
PR 28-OCT-1998; 98US-0106039.
PR 28-OCT-1998; 98US-0106042.
PR 28-OCT-1998; 98US-0106044.
PR 29-OCT-1998; 98US-0106072.
PR 29-OCT-1998; 98US-0106073.
PR 29-OCT-1998; 98US-0106074.
PR 29-OCT-1998; 98US-0106087.
PR 02-NOV-1998; 98US-0106587.
PR 02-NOV-1998; 98US-0106588.
PR 02-NOV-1998; 98US-0107034.
PR 02-NOV-1998; 98US-0107035.
XX
XX (CONN-) CONNAUGHT LAB LTD.
PA
P1 Murdin AD, Oomen RP, Wang J;
XX
XX WPI: 2000-350688/30.
DR
XX N-PSDB: A28710.
XX
XX Chlamydia antigens and the proteins they encode, useful for
PT vaccinating against Chlamydia infections that affect the respiratory
PT tract
XX
XX Claim 13: Fig 21: 226pp: English.
XX
XX The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
CC C. pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins
CC in samples (e.g. via enzyme linked immunosorbent assay (ELISA)).
CC
XX
XX Sequence 1000 AA:

Query Match 20.8%; Score 1057.5; DB 21: Length 1000;
Best Local Similarity 30.5%; Pred. No. 5.4e-70;
Matches 302; Conservative 166; Mismatches 446; Indels 77; Gaps 28;
OY 6 FFELIGSLSLAREVSRIFLAFNSVPD-----PTKESLSNK-----ISL 46
AC
XX
XX 56 flllygnftacmfgmcpavyslqdslekfalerdeefrslpdlstltgfsplttf 115
DT
XX 47 TGDTHNLT-NCYLDNLRYIIAIILOKTPMEGAAYITTDVLSFFDQKBEIYPAKNUTPESG 105
DE

Db 116 vgnrhssqddvlylnyksidnllllwsagavscnftl--lsnvedahfisknlaigt 173
OY 106 GAIGYASPNSPVEIRDTIGPIEFENNTCCRPFTSSNPAAVANKIREGAIHQ-NLYIN 164
Db 174 gaaacqg----actlktkrpdlffsnrgln-----nastggetrghaiaacngdfcis 222
OY 165 HNHVDVGFPMKNEFSYRGCAISTANTFVVSQSCFLFMDNICIQNTAGKGAIACTSN 224
Db 223 qnqgflfyfnsvnmvggalstngchrlsqnraplllffn-----ntapsggallrsnt 276
OY 225 SPESNCCCLFPIYNNACGAGAFSPF-CSLGNRCNIIYFVNNRCRKNVETASSESDGA 283
Db 277 tisdntriyltkmngnnggalqtsvvalknsgsvlfnmtalsg--slnsgrnsgga 334
OY 284 IKVTRLDVTGNRGRIFESDNTTKNYGAIYAPVTVLDNGPFTYFINNIANKGAIYID 343
Db 335 l-yctnlsiddnpgyllfnmnyctrdggaicqflltknsglhyftm-qgnwggaimgl 392
OY 344 GTSNSKISADRHAIIFENIYVNTNANGTSTANPPRMAITVASSGCEILLGAGSSON 403
Db 393 qdstclllaegnlafqnevflltffg-----fynaibhcupns-nlqlgankgyl 441
OY 404 LIFVDPFEVSNAGVS-VSFNKEADOTGSVVSFGATVNSADFHQRMLQRTAPLTLNMGF 462
Db 442 tafldpbehqhtnplrlfnpnanhgtllfssaylpaasdyennflssknrtselngv 501
OY 463 LCIEDHQAOLTVNRFTOTGVVSLGAVLSCYKKNAGNSAS-NASITLKHIGLNLSTLK 521
Db 502 lsiedragwqfkyftqkqgllkighaaslatansetpsvsgvllmnaimgl 561
OY 522 SGAEIPILWEPTNNSNNYADTAATFSLSDVKLSLIDYGNSPESYDNLTHALSSQML 581
Db 562 kg-kapltwirplqssapftednnpdltslsg-pllllneenrdpydsldseplqnlll 619
OY 582 SISEASDNOLASDDMDQFGLN-VPHYGNQGLMTGCMATODPEPASATTTDPOKANRFH 640
Db 620 slsdvlarhlnldhpselinatehygqslwspwvett--lfttnasi--etlanly 674
OY 641 RTLLLTWLPAGVSPSPKRSPLIANTLWGNMLLATESLKSNAELTPSD--HPFGITGG 698
Db 675 ralyanweplygkvnpeygdlatplwqsfthmtllrsytrngdslerfleiqla 734
OY 699 LGMWYODPRENHDPGPHRRSSGYS--AGMIAGQTHFSLFKTSQTYTKLERYAKNNVSSK 756
Db 735 dglfvhqnslpgabqftrqstcyslqasetslbnklsldglaqftrtrkeisnnyvsh 794
OY 757 N-YSCGEMLFSLQEGFLTKLVGLVSYGDHNCHEFTTQGENLTSQTFPSQTMGCAVF 814
Db 795 ntvaslyvelpw-fgeatatsla-ygygdhlyharylrhkn-raegtcyshtlaaig 851
OY 815 FDLPMKPPGSGNHTLTAPLALGIYSSLSHTEVGAVRSPSTKPLNLNVLPVGVGSE 874
Db 852 csfwqgqsyhl--spfvgaiairshqtafeeygdnpkrkvtvsgpfpynlclpigiqlw 909
OY 875 MNATQRQAWTVELAYOPVLYROEPGIATOLLASKGIWFGSGSPSRAMSVKISQOTOP 934
Db 910 qskfhvprewlelsyqpvlyqnpqigvltllaasgsdvlghmyvrnalykvhnqal 969
OY 935 LSWLTLHRYHGFYSSSTFCVYLVNGELALRF 965
Db 970 frsldlfidygsvsvsstshlqagstlkt 1000

RESULT 8
ID Y95551 standard; Protein: 963 AA.
XX Y95551:
AC
XX
XX 10-OCT-2000 (first entry)
DT
XX Chlamydia pneumoniae antigen CPN100624 RY-64.
DE

PD 04-MAY-2000.
 XX 28-OCT-1999; 99WO-CA00992.
 XX 28-OCT-1998; 98US-0106034.
 PR 28-OCT-1998; 98US-0106039.
 PR 28-OCT-1998; 98US-0106042.
 PR 28-OCT-1998; 98US-0106044.
 PR 29-OCT-1998; 98US-0106072.
 PR 29-OCT-1998; 98US-0106073.
 PR 29-OCT-1998; 98US-0106074.
 PR 29-OCT-1998; 98US-0106087.
 PR 02-NOV-1998; 98US-0106587.
 PR 02-NOV-1998; 98US-0106588.
 PR 02-NOV-1998; 98US-0107034.
 PR 02-NOV-1998; 98US-0107035.
 XX (CONN-) CONNAUGHT LAB LTD.
 XX Murdin AD, Oomen RP, Wang J:
 PI WPI: 2000-350688/30.
 DR N-PSDB: A28708, A28709.
 XX Chlamydia antigens and the proteins they encode, useful for
 PT vaccinating against Chlamydia infections that affect the respiratory
 PT tract
 PS Claim 13: Fig 19; 226pp: English.
 XX The nucleic acids may be used for the recombinant production of the
 CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
 CC recombinant DNA methodologies. The polypeptides may then be used to
 CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
 CC C. pneumoniae, are pathogens responsible for upper respiratory tract
 CC infections such as community acquired pneumonia, acute respiratory
 CC disease and bronchitis and may be implicated in atherosclerotic changes
 CC and asthma. The nucleic acids may also be used as probes for detecting
 CC the presence of Chlamydia nucleic acids in samples (and therefore
 CC diagnose infections) and the proteins may be used as antigens for the
 CC production of antibodies that may be used to detect Chlamydia proteins
 CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
 XX Sequence 954 AA:

Query Match 18.6%; Score 947.5; DB 21; Length 954;
 Best Local Similarity 29.1%; Pred. No. 7.8e-62;
 Matches 281; Conservative 166; Mismatches 418; Indels 101; Gaps 34;

OY 35 PKRESLSKISLTGDTNLT-----NCYLDNLRLILAILOKTPNEGAAVTTDTYL 84
 DB 55 plldtln---mtpyshratlfgvrddtngdivldhqnieswfeifsgdgaalscks-1 110
 OY 85 SEFDPOKEGIYFAKNLTPESSGAIYASPSPTVEIRDIGVIFENNCCAPFTSSNP 144
 DB 111 altnct-kingillinsfalkkragam-yvdgn---foisenhgsilltsgnl-----stpn 158
 OY 145 AA--VNKIREGCAIHAONLYINHNHDVYGFMKNFYSVROGAISTANTFVSENOGCLPM 202
 DB 159 asnfadctcgvavcskvntlekngtlayflnnkaksaggaalnlndkdtlpgclff- 217
 OY 203 DNICIGTNTAG--KGAIAIAGTSNFSFNCDLFNNAACCGAGAI--FSPICSLGNRCN 259
 DB 218 -----nnaagtaggaifanacr-iennsqilyflnngsiggalirvhecllknltgs 270
 OY 260 IYFVNNRCFKNVEIASSEASOGCAIKVTRLDVGNRGRIFFSDGNTKNGAIVAPVYT 319
 DB 271 vlfmnn--lameadisahnssggaalycis-csikdnpylaadtmhaardgaalictqst 327
 OY 320 LVDNCPYTFINNIAKNGGAIYIDGTSNSKISADRHAIIFENIYTNVNTANGTSTSANP 379
 DB 328 lqdsqpyvftmn-qgtwgalmlrqdgactllfadgdgdliflynnrnfkdtsfn--hvsync 384

OY 380 PRRNATVASSGCEILLGAGSSONLIFYDPI----EVSNAQSVSFNKKADOTGSAVFSG 435
 DB 385 trnvsltyvasgq-----hsatfydbllqrytlqns--lqkfnnpnphllygtllss 433
 OY 436 -----ATVNSADF--HORNLOKTRPAPLTSNGFLCIEBHAOLTVRRFGTGGVSLGAG 488
 DB 434 tyipdstlsrddiffnfrn-----hlyngltaladbraekwykfkqfqqglrlr]gstr 486
 OY 489 AVLSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEPLWVPTNNSNNTADTAATF 548
 DB 487 avfstldeeqsssvsgvnlhnlalnlpoll-gnvapklrlprgssapysecdnpll 545
 OY 549 SLSDVKLSLIDYGNSPYESTDLTHALSQPMLSISEASDNDLRSDMDPSCGLN-VPIHG 607
 DB 546 nlsq-plsllddenldpytdadlaqplaevpillylldvtakhlnthdnfyepgnttqhyg 604
 OY 608 WQGLMTWGWAKTQDPEPPASATITDQKANRFRRLTLTLWLPAGYVPSPKHRSPLIANTL 667
 DB 605 yqgvswpywieclitstdts-----edlvnclhrqlgydwpclgykvnpdenxgdalalsaf 659
 OY 668 W---GNMLATESLKNSEALTPSDHPFWGITGGLGMVYQDPRENHPGFHMRSSGSAG 724
 DB 660 wgsfnlfcaltlryqgqglapt-----asgeatrlfwhqsmndakylfmeatqyslg 713
 OY 725 MIA--GQITFSLKFSQITKLNERYAKNNVSSKNVSCGEMLFS-LOEGFLTLKLVGLY 781
 DB 714 tsnlasnhsfgvnfsqjlsnlyeshsdnsvashettvalqlnpwlqgerfatsasla-y 772
 OY 782 SYGDNCHHFTYTOGEN--LTSGTFRSOTMGAVFEDLPMKRFGSHILTAFLALGTY 839
 DB 773 sygn---nhlkasgysgkqteggkycstllgaalscsislg-wrsrlplfcplqalavr 828
 OY 840 SLSHTEVGAVPRESFTKTPLINLVPIGVGSEFMNATQROAVTVELAYQVLYROCP 899
 DB 829 snqtlfgesgdarkfsvhkpnylntlpiglsaweskfrlplynhielaypvlvqgnp 888
 OY 900 GIATQLLASKGIWFGSGSSSHNHSYKISQOTOPSLWTLTFHOYHNGFYSSSTFCNYLNG 959
 DB 889 einvsiesgsswllsgtlatrnaiafkgrnqllfipklsvfldygvssstcthyha 948
 OY 960 EIALRF 965
 DB 949 gtlfkf 954

RESULT 10
 Y92831
 ID Y92831 standard; Protein: 788 AA.
 XX Y92831;
 XX 29-AUG-2000 (first entry)
 DE C. pneumoniae CPN100662 processed antigen.
 XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
 KW anti-arteriosclerotic; vaccine.
 OS Chlamydia pneumoniae.
 PM WO200024765-A2.
 PD 04-MAY-2000.
 XX 28-OCT-1999; 99WO-CA00992.
 PF 28-OCT-1998; 98US-0106034.
 PR 28-OCT-1998; 98US-0106039.
 PR 28-OCT-1998; 98US-0106042.
 PR 28-OCT-1998; 98US-0106044.
 PR 29-OCT-1998; 98US-0106072.
 PR 29-OCT-1998; 98US-0106073.

```
Query Match 14.3% Score 727; DB 20; Length 822;
Best Local Similarity 28.1% Pred. No. 1.6e-45;
Matches 234; Conservative 140; Mismatches 355; Indels 106; Gaps 35;

OY 35 PTKESLSNKSISLTGDTNHLT-----NCYLNDLRYLIALLOKTPNEGAAVTTDVL 84
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 48 plidltln---mtyshratlfgvrdtndqdividhngmsiesfentfgdgalscks-1 103
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 85 SFEDTQKEGIIYFAKNLIDREGGALIGVSPNSPTVEIRDITGVIFENNTCCRPFTSSNPN 144
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 104 alntc-kngillfnstakragam-yvngn---fdlsehgslfsgnl-----sfpn 151
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 145 AA-VNKRREGAIIHAQNLVYINHNHDVGFPMKNFSYVGAISTANTVVSNSQCFLEM 202
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 152 asnfadctggavlescknvtlskngtrayflnkkakssgqalgaalnlkdhltgpltf- 210
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 203 DNICLOTWTAG--KGAIVAGTSNFSFESNCDLFFINNACGAGAI-FSPICSLTGNGRN 259
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 211 -----maagxtaggalifanacr-iennsqpiylfmgsglvggalrvhgecltkngs 263
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 260 IVFYNNRCFKAVETASSASDGAIKVTRLDVYGNRGRIFFSDNITRNNGAIVAPVVT 319
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 264 vlfnnn--fameadisahnssggaicis-csikdnpjlaafndntaardgaictgalt 320
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 320 LVDNGPTTFINNANNKGAIVYIDGTSNKSISADRHAIIFNNENITVNTNANGTSTSANP 379
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 321 lqdgspvylfnn-qgltwgalmllrgdgacclfdagddllfynrhfkdlfn--hvsync 377
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 380 PRNRAITVASSSGEILLGAGSSQNLIFYPDI---EVSNAQSVSFNKEADDTGSVFSG 435
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 378 trnsrltvgagq-----hsatfydplllgyltigns--iqdfnpnpehlgilits 426
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 436 A-----TVNSADF--HQRNLOTKTPAPLTLNSGFLCEDHAQLTVNRPQTGTGVVSLNG 488
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 427 ayldptstrddffishfrn-----higlyngtialdraekwykfdqfgsltrlgsr 479
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 489 AVLSGCKKAGNASNASNASTIKHIGLNLSTILKSGAEIPLMLVEPTNNSNNTATATATF 548
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 480 avfstldeegsssvgavlinlnalnlpalsl-gnrvapklwlrptgsapysednpll 538
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 549 SLSDVKLSLIDDYGNSPYESTDTHALSSQPMLSISEASDQLRDEDDEGLN-VPHYG 607
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 539 nlsq-plsllddenldpdtadlagpiaeyrplyldvtakhindntypeeglnctqhyg 597
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 608 WQGLTWGMAKTODEPARASATITDPOKANRFHRTLLTLWLPAGVSPKRRHSPLIANTL 667
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 598 ygvvswpywlaetltsdts-----edlvnclhrqlygdwlrptgkvnpenkgdialsaf 652
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 668 W---GNMLLATESLUNKSNALTPSDHFWGIRGCGGGMVYDDPRNRHNGCFHMRSSGYASG 724
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 653 wqslnllactlrytqgqgiapt-----asgeatrlivhngsnndakgflmneaigyalsg 706
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 725 MIA--GQHTFSLKFSQTYTKLINERYAKNNVSKNYSCQGEWLS-LOEGFLLTKLVGY 781
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 707 tctnashshsfynfsqflnlyeshedsnvashtltvalqinmpwldertstssla-y 765
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

OY 782 SYGDHNCHEFTQGEN--LTSQGFRRSQTM-GGAVFPLPKMFGSTHILAP 831
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 766 sysn--hthkasgysskltqegkcykstllrgsllsl-----stmaaltcp 810
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 12
Y16738 ID Y16738 standard; Protein: 1013 AA.
XX AC Y16738:
XX 21-JUL-1999 (first entry)
XX C. trachomatis F serovar HMW protein.
XX Chlamydia; high molecular weight protein; HMW protein; urethritis;
```

```
KW bacterial infection; conjunctivitis; lymphogranuloma venereum; LGV;
KW cervicitis; epididymitis; endometritis; pelvic inflammatory disease;
KW PID; salpingitis; tubal occlusion; infertility; cervical cancer;
KW arteriosclerosis; atherosclerosis.
OS Chlamydia trachomatis.
PN W09917741-A1.
XX 15-APR-1999.
XX 01-OCT-1998; 98WO-US20737.
XX 02-OCT-1997; 97US-0942596.
XX (ANTE-) ANTEX BIOLOGICS INC.
XX Jackson JW, Pace JL;
XX WPI: 1999-287659/24.
XX
XX New Chlamydia protein useful for treating conjunctivitis, urethritis
XX and cervical cancer
XX
XX Claim 4; Page 119-123; 141pp; English.
XX
XX The invention relates to an isolated Chlamydia species high molecular
XX weight (HMW) protein having an apparent mol. wt. of 105-115 kD as
XX determined by SDS-PAGE. The HMW proteins and nucleic acid molecules can
XX be used for preventing, treating or ameliorating a disorder related to
XX Chlamydia e.g. bacterial infection, conjunctivitis, urethritis,
XX lymphogranuloma venereum (LGV), cervicitis, epididymitis, endometritis,
XX pelvic inflammatory disease (PID), salpingitis, tubal occlusion, cervical
XX cancer, infertility, arteriosclerosis and atherosclerosis. The products
XX can also be used for detection and diagnosis. The present sequence
XX represents a C. trachomatis HMW protein.
XX
XX Sequence 1013 AA:
XX
XX Query Match 12.8% Score 651.5; DB 20; Length 1013;
XX Best Local Similarity 25.6% Pred. No. 9.2e-40;
XX Matches 260; Conservative 152; Mismatches 389; Indels 213; Gaps 43;
XX
XX 39 SLSNKSISLTGDTNHLFNCYLNDLRYLIALLOKTPNEGAAVTTDYLSEFDTQKEGYFAK 98
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 84 nllegfvlgrghslt---fenlr-----tsungals-----daanglf--- 121
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 99 NLTPESGAIGVASNPTVEIRDITGVIFENNTCCRPFTSSNP-NAAV----- 147
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 122 --tleqfkelstfscnslavt---paatlmgsgqlptltstpsngltlysktdl111n 174
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 148 -----NKIR-EGGAIIHAQNLVYINHNHDVGFPMKNFSYVGAISTANTVVSNSQ 198
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 175 nekftfynlvsogdgtldaksilvtqgisklcvfgentadagdgacqvvtfsamnap 234
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 199 FLEMDNICIQTNTAG-KGAIIYA-----GTSNSEE-----SNDCLEFFINNACG 243
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 225 laffianv-----agvrvgiaavqdgqgvsststedprvstfntlaveldgvarvg 288
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 244 GAIFPICSILTCNCRNIVFYNN--RCFKN-----VETASSEASDGA 284
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 289 991ys-----ygnvaflnngklflnvaaspyiaaegplngqaastdnyddga1 340
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 285 KVTFLDVTGN-----RGRIFFSDNITRNNGAIVAPVTVLVNCPYTFINNTAN 334
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 341 fckngagaagsnsgsvsfddgevgvffssnvaagkggalikkslsvancgpyvflgnlan 400
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 335 NKGAIVYIDGTSNKSISADRAHAIIFENI---VTNVTNANGTSTSANPPRRNAITVASS 390.
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 401 d-ggalilygeegelsadylgdgnlkrtakenaadvngvtvs-----qalsmgsg 454
XX |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
XX 391 SGELLGAGSSQNLIFYPFIVSN-----AGVSVSFNKEADDTGSVFGAIVNSADF 443
```

QY 693 GITGCGLCGMVQODRENNPCHMRSSGYS--AGMIAGOTHPFSLKFSQTYTLYNBRYAK 750
 Db 641 -----gldfllnedkkgngqrsyrhssagyalgggffltaseenflnfafcdqf-----gydk 650
 QY 751 NNWSSKNYS--CQGMFLF-SLQEGFLTLTVG-----LYSYGDHNCNHFY 793
 Db 691 dhlvaknlthvyagamsyrlhbskltleklsngsdlpfvfnarlayghtldnmmtkyt 750
 QY 794 QGENTLSQGTSTGTM---GGAVFFDLPMKPPG-----STHILTPFLGALGITYSSLH 844
 Db 751 gyspv--kxswgndafigiecgagl---pvwasgrswdth---tprflnlemiyahqnd 801
 QY 845 FFEVGAAPRSFTKTRPLVLVLPIDVKGSSFMNATORQAWYVELAOPVLYRQEPRLAQ 904
 Db 802 fkenlegstgisg-dflnlavpvglk--fekfsdk--stydslslayvpdlrddpccctt 857
 QY 905 LLASKGIWFGSGSPSSRHMSYKISQOTRPLSWLTLHFQYHNGFYSSSTPCNYLNGELIALR 964
 Db 858 lmvsgdswstcgtsisrgallvrag-----nhhaf--asnfefisqfewelr 902

 RESULT 14
 B13633
 ID B13633 standard; Protein; 982 AA.
 XX
 AC B13633:
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE C. trachomatis pmpg gene protein.
 XX
 KW Chlamydial infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 XX
 OS Chlamydia trachomatis.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 981
 FT FT
 XX
 PN WO200034463-A2.
 XX
 PD 15-JUN-2000.
 XX
 PF 08-DEC-1999; 99WO-US29012.
 XX
 PR 08-DEC-1998; 98US-0208277.
 PR 08-APR-1998; 98US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skeiky YAM, Fling SP, Jen S, Stromberg EJ;
 XX
 DR WPI: 2000-431303/37.
 PT
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 XX amino acid sequence encoded by polynucleotide sequence -
 XX
 PS Claim 2: Pages 181-184; 256pp; English.
 XX
 CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can

Query Match	12.7%	Score 647	DB 21	Length 982
Best Local Similarity	25.4%	Pred. No. 1.9e-39		
Matches	257	Conservative 156	Mismatches 388	Indels 212
				Gaps 43
XX	Sequence	982 AA:		
CC	led to pleuritic inflammatory disease (PID), resulting in tubal obstruction			
CC	and infertility. Trachoma due to ocular infection with <i>C. trachomatis</i> is			
CC	the leading cause of preventable blindness worldwide. <i>C. pneumonia</i> is a			
CC	major cause of acute respiratory tract infections in humans and is also			
CC	thought to play a role in the pathogenesis of atherosclerosis and			
CC	coronary heart disease. The present sequence is a protein isolated in the			
CC	present invention.			
QY	39	SLSNKISLTGDDHNHNTLCYLDNLRVIALAQOTPRNGCAAVTTIDYLSFFDTQKEGIYFAK	98	
DB	54	nlgsgfyvlgryhstl--fenlr-----tstngaals-----nsaadgfl---91		
QY	99	NLTPESSGALIGASPNSPTEVIRDTIGVIFENNTCCRPFTSSNP--NAAV-----147		
DB	92	--tiegfkelsfscnslsllavi-----paatlnkgsqrpttstetngtisktdllln144		
QY	148	-----NKIR--EGGATHAQNLVLIHNHNDVYGFPMKNFSTVRCGALISTANTFVVENOSC198		
DB	145	nekfsgfynlvsqdgdaiaakslvtvgysklevfgentagdgagacqvvtlsfiamenap204		
QY	199	FLFMNICTQIOTYAG--KGGAIYA-----GTSNFE-----SNMCDLFFINNAACAG243		
DB	205	iafvnrv-----agrvgggiaavdgqdgqvssstetcdprvysfrntavefignvayrg258		
QY	244	GAIFSPICSLTGNKGNIVFYNN--RCFKN-----VETASSEASDCAI284		
DB	259	g9iys-----ygnvatflnngklflnnvaspvyiaakqptsgqasntsmnygdgal310		
QY	265	KYTTFLDITGN-----RCRIFFSDNITKNYGATYAPVTVLVNDNGTYYINNIANN335		
DB	311	fckngagagsgnsgsvsfqegvvtffssnvaagkggaayakklsvancaqpvqflrnlrand370		
QY	336	KCGAIYIDQTSNKSIAADRAHIIIFENI---VTNTNNGTSTSNANPRRNITVAASS391		
DB	371	--ggaalyigesgelslsadqgdllfdgnlkrtakenaadvngvtvss-----qaismgsg9424		
QY	392	GEILLGAGSSQNLIFYDPIEVSN-----AGVSVPFNKEADQTSVSGATVNSADPH444		
DB	425	klttlrakagbqllfndpremanngnqpqsgskllkldngeytlgdlvfn---gssltly481		
QY	445	QHNLOTKTPAPRLTSLNGFLICIEDHAOLTVNRFOTQCGVSVSLGCAVLSCYKKNAGCNS--A502		
DB	482	gn-----vtiegqrvltrekaklsvnsislqtgslslymeagstldfvtcpqppqpa532		
QY	503	SNASTIKHIGINLSLILKSGAEIRPLMVEPRN--NSNNTATATATFSLSDVKLS---L557		
DB	533	aqqlitflnlnhslsllanna---vtprtpptpdpasipavagst--tegsvtlsgpflf587		
QY	558	IDVGNSPPESTD-----LTHALSSQPMILSISEASDNOLRSDDMDFCLNPNHYG607		
DB	588	fedldtdetvdydwsgnqklnvklkqigtkr--panapsdill-----gnempkyg637		
QY	608	WQGLMTWGMAKTQDPEPASSATITDFOKANRPHRTLILTLVACGYVSPKRNRPILANTL667		
DB	638	yggswklaw-----dpteanngpyclkatwcktgygpprevaslvpnsl682		
QY	668	WGNMLLATSLKNSALELPSPDHPFW--GIIGGCGGMVNVODPRENHGCFHNRRSGYSGMI726		
DB	683	wgs--lldlrshaisaiqasvdgrsycrglwwsgsnfflyhndaiggytyvisgyslig--739		
QY	727	AOQTHFSLKFSQTYTKLBERYAKANNV--SSKNVSCOGEMLFSLQEGFLTLTKLVGLATSYG784		
DB	740	--ansyfgssmiflafevfygr--skdyvvcsmhactlgsyvlstqga-----lqgsylf92		
QY	785	DHNCHEHYTQG--ENLTSGCTFNSQT-----MGAAVFEDLPKMPFGSTHILT--APFL833		

AC Y16737:
 XX
 DT -21-JUL-1999 (first entry)
 XX
 DE C. trachomatis B serovar HMW protein.
 XX
 KM Chlamydia: high molecular weight protein; HMW protein; urethritis;
 KM bacterial infection; conjunctivitis; lymphogranuloma venereum; LGV;
 KM cervicitis; epididymitis; endometritis; pelvic inflammatory disease;
 KM PID; salpingitis; tubal occlusion; infertility; cervical cancer;
 KM arteriosclerosis; atherosclerosis.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9917741-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 01-OCT-1998; 98WO-US20737.
 XX
 PR 02-OCT-1997; 97US-0942596.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Jackson JW, Pace JL;
 DR WPI; 1999-287659/24.
 XX
 PT New Chlamydia protein useful for treating conjunctivitis, urethritis
 PT and cervical cancer.
 XX
 PS Claim 4: Page 115-119; 141pp; English.
 XX
 CC The invention relates to an isolated Chlamydia species high molecular
 CC weight (HMW) protein having an apparent mol. wt. of 105-115 kD as
 CC determined by SDS-PAGE. The HMW proteins and nucleic acid molecules can
 CC be used for preventing, treating or ameliorating a disorder related to
 CC Chlamydia e.g. bacterial infection, conjunctivitis, urethritis,
 CC lymphogranuloma venereum (LGV), cervicitis, epididymitis, endometritis,
 CC pelvic inflammatory disease (PID), salpingitis, tubal occlusion, cervical
 CC cancer, infertility, arteriosclerosis and atherosclerosis. The products
 CC can also be used for detection and diagnosis. The present sequence
 CC represents a C. trachomatis HMW protein.
 CC
 SQ Sequence 1013 AA:

Query Match 12.7% Score 646.5; DB 20; Length 1013;
 Best Local Similarity 25.5% Pred. No. 2,2e-39;
 Matches 259; Conservative 154; Mismatches 388; Indels 213; Gaps 44;
 OY 39 SLSENKISLTGDTNLTNCLYDLNRYLAILQKTPNEGAAVTITDYLSFPDQEGIFYFAK 98
 DB 84 nllgstfvlgrphslc---fentlr-----tstngaaals-----dsanglft--- 121
 OY 99 NLTPESGGAIGVSPNSPYEIRDITGPVFNNTCCRPSTSNP-NAAV----- 147
 DB 122 -tiegfkelfstfncn-pllavl---paatlngsqtpstltpstpsngtlysktdlllln 174
 OY 148 -----NKIR-EGCAIHAONLYINNHHDVGFMKNFYSVRCGALSTANTFVVSFNQSC 198
 DB 175 nekfstfysnsvsgdgaidakslvtvgisklvtfgentlaqdgagayvvsfamaeap 234
 OY 199 FLEMDNICIQNTAG-KGGAIVA-----GTSNSFE-----SNNCDFEFINNACAG 243
 DB 235 latvanv-----agvtyggiaavdgqgvssstsedpovvsftrltavefdgnvarvg 288
 OY 244 GAIFSPICSLTGNRGNIVFNN--RCFKN-----VETASSEASDCAI 284
 DB 289 gqlyfs-----ygnvalfnngkllflnnvaspvyaieeqptngqansdsndygdgaal 340
 OY 285 KVTTRILDVNTGN-----RGRIFPSDNTKKNYGAIYAPVVTLVNDGPTVFINNIAN 334

DB 341 fckngagaagsnngsvsfedgqvvtffssnvaagkggaljyakklsvanegpyqlgnian 400
 OY 335 NKGGAIYIDGTSNKSADRHAIIENENI----VNTVNTANGTSTSANPPRRNATVASS 390
 DB 401 d-ggaalygeesgealsisadgdmfIdgnlktakenaadngvltvss-----qalsmgsg 454
 OY 391 SCGILLIACGSSONLIIFYDIEVSN-----AGVVSFFKADQGTGSVVFSGATVVSAPF 443
 DB 455 gklttlrtaaqhllndplmangnngnpgagseplklndegvygdvlf--angnstly 512
 OY 444 HQRNLOTKTPADLTLSNGFLICIEDHAOLTVNRFOTQGVVSLGNGAIVLSCYKNGAGNS-- 501
 DB 513 gn-----vtlegrvlvirekalsvnsisqggslymeagstltfepqpqgpp 562
 OY 502 ASNASITLKHIGLUNSSILKSGAEIPLWVEPN--NSNNYTDATNATSLSDVKIS--- 556
 DB 563 aangstlllenlnlslslanna-----vlnpplnpgadshpavlsqst-laagsvltsgpl 617
 OY 557 LIDDPGNSPYESTD-----LTHALSSOPMISISEASDNLKSDMDPSGLWVPHY 606
 DB 618 ftedldtdaydydwlgnsqklavlkqqlcp--panapsdltl-----gnempky 667
 OY 607 GMOGLWTGMAKTODEPPASSATITDPKANRPHRTLLTLWLPAGYVPSPKHRSPLIANT 666
 DB 668 gyqgsklaw-----dntcanngpylkatwktkytnppervasiypns 712
 OY 667 LMGNMLATESLKNSELPSPDHPFW-GITGGGLGMVYODPRENHPGHMSSGTSAGM 725
 DB 713 lwgs-llldlrsalsalqaesvdgrfscrglsvsgvsnffhldrdalqgyrytysgyslg 770
 OY 726 IACQHTFSLKFSQTYTKLINERAKNNV--SSKNVSCOCBMLFSLDGGFLTLVLYLSY 783
 DB 771 --ansyfgssmglatlewfgr-skdyvvcrrnhmclsisvylstqda-----lgsyvf 822
 OY 784 GDHNCHEFTYTOG-ENLTSGTFRSQT-----MCGAVFEFDLPMPKPFGSTHILT--APF 832
 DB 823 gdfafraayfyfgnkmktyltfaeesdvcwncnclvgeiygjpivltpskylinelrpf 882
 OY 833 LGAIGYSSLSHTEVGAVPRSFSTPLINLVLPVIGVKGFSFNATQRPQAMTVELAYOP 892
 DB 883 vqetfayadhesltleegdqaralfrs-ghlmnlsvpvkxfidrcsst-hpnkysfmgaylc 940
 OY 893 VLVKRGECIATOLLASKGIWFGSCSPSSRRHMSVKISQOQPSLTIHFOYHG 946
 DB 941 dayrltslqqltllshqetwtdafnarngvltvrgmzas-----ltsnlevy 990

RESULT 17
 Y99843
 ID Y99843 standard; Protein: 925 AA.
 AC Y99843:
 XX
 DT 15-SEP-2000 (first entry)
 XX
 DE Chlamydia pneumoniae processed 98 kDa outer membrane protein CPN100640.
 XX
 KM Chlamydia: 98 kDa outer membrane protein; antigen; immunogen; infection;
 KM vaccine; antibacterial; community acquired pneumonia; bronchitis;
 KM sinusitis; acute respiratory disease; upper respiratory tract disease;
 KM asthma; atherosclerosis.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO200032784-A1.
 XX
 PD 08-JUN-2000.
 XX
 PF 01-DEC-1999; 99WO-CA01148.
 XX
 PR 01-DEC-1998; 98US-0110439.
 PR 03-MAY-1999; 99US-0132272.

CC community acquired pneumonia and upper respiratory tract symptoms and
CC diseases, including bronchitis and sinusitis. It also has an association
CC with atherosclerosis and asthma. The 98 kDa outer membrane protein is a
CC C. pneumoniae-specific antigen which can confer immune protection against
CC chlamydial infection. The nucleotide sequence encoding the protein or the
CC protein itself may be administered as a vaccine to prevent or treat
CC infection and they may also be used to diagnose infection. The gene
CC encoding CPN100640 was amplified from Chlamydia pneumoniae genomic DNA
CC by PCR.
xx
Sequence 936 AA:

[illegible]

Query Match	12.7%;	Score 645;	DB 21;	length 936;
Best Local Similarity	27.0%;	Pred. NO. 2.5e-39;		
Matches 262;	Conservative 146;	Mismatches 392;	Indels 172;	Gaps 44

RESULT	19
Y16735	
ID	Y16735 standard; Protein; 1012 AA

DT 21-JUL-1999 (first entry)
XX
DE
XX C. trachomatis LGV L2 HMW protein.
KW Chlamydia: high molecular weight protein; HMW protein; urethritis;
KW bacterial infection; conjunctivitis; lymphogranuloma venereum; LGV;
KW cervicitis; epididymitis; endometritis; pelvic inflammatory disease
KW PID; salpingitis; tubal occlusion; infertility; cervical cancer;
KW atherosclerosis; atheroclerosis.

Chlamydia trachomatis

PN W09917741-A1.

PD 15-APR-1999.

PF 01-OCT-1998; 98WO-US20737.

PR 02-OCT-1997; 97US-0942596.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson JW, Pace JL:

WP1: 1999-287659/24

DR N-PsDB; Xb0539.
XX

PT New Chlamydia protein useful for treating conjunctivitis, urethritis and cervical cancer

XX
XX
C
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840

CC The Involution relates to an isolated Chlamydia species high molecular
CC weight (Hmw) protein having an apparent mol. wt. of 105-115 kD as
CC determined by SDS-PAGE. The Hmw proteins and nucleic acid molecules can
CC be used for preventing, treating or ameliorating a disorder related to
CC Chlamydia e.g. bacterial infection, conjunctivitis, urethritis,
CC lymphogranuloma venereum (LGV), cervicitis, epididymitis, endometritis,
CC pelvic inflammatory disease (PID), salpingitis, tubal occlusion, cervical
CC cancer, infertility, arteriosclerosis and atherosclerosis. The products
CC can also be used for detection and diagnosis. The present sequence
CC represents a c. trachomatis Hmw protein.

Sequence 1012 AA;

Query Match	12.6%	Score 642;	DB 20;	Length 1012;
Best Local Similarity	25.3%	Pred. No. 4.7e-39;		
Matches 259;	Conservative 155;	Mismatches 376;	Indels 234;	Gaps 45;

```

0Y 39 SLSNKISLTGDTHTLNCYLDNLRYLAILQKTPNEGAANTITDYLSFFDQKEGIYFAK 98
      :|  :: | | | |  ::| |  | | | |  ::| |  :| | :
Db 84 nlgsftvgyrghst--fehir-----stngaals-----nsaadylf--- 121

```

D	b		258	--dgnkagampdillltsqne-slnflm-----lagnsgeai-ytkkiavlssg	302
O	y		296	RGRLEFSQNTIKXNKGAVIAPVVTLVDNCPYTFINNIANKKGAITYIGCTSKRSADRH	355
D	b		303	rgvvlfsnhaan-----alpkgaatalldseislsadlg	338
O	y		356	AIFENENIVTNVANQGTSTANP--RNAITVASSGEILIGASSQNIFYPIEVIS	413
D	b		339	nlf-----egntstgspsavtnaldasnakfmlratryknkvfydpitss	389
O	y		414	NAGVSVPNKREAD-----QTGSVSSGATVNSADPHQ-RNLQTTPRPLPLSNGLFICIE	466
D	b		330	gatcklslnk-adagsnlyegyivlsegekleeelkkpdknlksftgavlaagaavljk	448
O	y		467	DHAOLTVNRFTQTGGVVSLDNCAGVALSCYNKNGASNSASNASTLKHIGHNLSTLSKSAEI	526
D	b		449	dgvlevantlltqegskvwmvd-----gttfesaagvlinglalnlsl-	493
O	y		527	PLWVEPTNNSNNTADTAFTSLDYKUS---LIDDYGSPVESTDLTHALSSQ---P	579
D	b		494	----dgtnk-----alikataaskvalasprlmvdaagny-ygh-----hnlsgqgyfv	538
O	y		580	MLSISESDNQSLDDMDFFSGLNV-P-HYMOGLMTMGMAKTYODPEPPASAITTPORANR	638
D	b		539	lleis-agctmtctdlpdrplntctbhyyggcgllvw-----ddatakt-----	583
O	y		639	FHRILLTLTWLPAGVYSPKHKRSPLIANTLMGN-----MLLATESTKNSAELTSPD	688
D	b		584	--knaeltlwcktygkpnpcegrplvpnsfwgfvdvatrsqslmdrstessisslnlws-	640
O	y		689	HPFMGITGGGLGMAMYVODPRENHIPGCHMRSSGY--AGNIAGQHTSFESLKFSQTYTKLNE	746
D	b		641	-----giadflhedqkngqrsyrshesagyalggfftasenfnfafcqjif----	686
O	y		747	RYAKNNYSKNYS--COGEELF-STOEGFLKLKLYG-----LYSYGDHNCH	789
D	b		687	gyddchlvahkanhlhyagamsyrhjgesktlakllsgnsdlprvfafarlayghtldmnt	746
O	y		790	HPYTQGENLTSOGTFRSQTM---GCAGVFDPUMKPFQ---STHILTAPFLCALIYS	840
D	b		747	tkyygyspv--ksgwgdalgiecgal---pwvasgriswldh---tpflndemlya	797
O	y		841	SLSHFTEVGAYPRSFEStKTPLINVLVPIGVKSFPMAATQROAMTVELATPOLYLROEPG	900
D	b		798	hqndfkmglegfrsfqe-dlfnlavpyvik--fekiedk-stydlaslayvdpdyindpg	853
O	y		901	IATOLLASKIMFGSSPSRHMAYSKKIKSIQOTQPSMWLTJHFHYHGFFSSSTFCNYLNGE	960
D	b		854	cettlmwsgswstcgtstlsrqallivrsg-----nhhaf--asnfevisqfe	898
O	y		961	IALLR 964	
D	b		899	velfr 902	
<hr/>					
RESULT 21					
ID	r35052				
ID	r35052 standard; Protein; 930 AA.				
X	x				
X	AC				
X	r35052;				
X					
D	T				
X	xx				
X	13-SEP-1999 (first entry)				
D	E				
X	Chlamydia pneumoniae surface exposed polypeptide.				
X	KM				
X	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis				
X	susceptible; purulent otitis media; erythema nodosum; pharyngitis;				
X	KW				
X	vaccine; neutralising epitope.				
X	xx				
X	OS				
X	Chlamydia pneumoniae.				
X	XX				
X	W09927105-A2.				
X	NN				

03-JUN-1999.

20-NOV-1998; 98WO-IB01890.

04-NOV-1998; 98US-0107078.

21-NOV-1997; 97FR-0014673.

(GEST) GENSET.

Griffais R;

WPI: 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Page 940-942; Disclosure: 191pp; English.

YJ4584-YJ35879 represent the proteins encoded by all the open reading frames in the complete genome (see XJ1990) of Chlamydia pneumoniae. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see YJ4584-YJ35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

Sequence 930 AA;

Query Match 12.2%; Score 620; DB 20; Length 930;
Best Local Similarity 27.1%; Pred. No. 1.8e-37;
Matches 252; Conservative 139; Mismatches 374; Indels 164; Gaps 40;

47 TGDTHNLTCYLDMRLRYLAILOKTPNEGAAT---ITDYSFPTDQKEGYFAKNLPPE 103
| | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
50 tada-ngnnylsagrvy-----ndaqgltalgccfctetgdltftkygsfstn-tvd 102
:
104 SCGAIGVASPMSPYVEIDT---IGPVIFENNCCRPFTSSNPAAVVKIREGGAIHQA 159
:
103 agsnagaasatladkallftgfsmstiaapgtvaagskstisagajnlndngtlfsfg 162
:
160 NLVIYNHNHDVVCFPMKMFESYVRGAISTANFTVSNSOCFLMDIQCOTNACK-GCAI 218
:
163 nvsneann-----ggal-taktlisngts-----siltssnakklgaI 203
:
219 YAGTSNFESENCDLFINN-ACCAGAI-FSPICSLTGNRGNIVFYNNRCKENETASS 276
:
204 yssaasaasigtglvfmmkxgetggagalgeaassitqn-sstifsgn-----tadt 255
:
277 EASGGGAIKV-----TTLDVDYGNRGRIFFESDNIRKNNGAIIAPVTVLVNDGPYPFIN 331
:
256 aagkggalcyckeyketptlitsgnks-ltfaensvtlggalaicahqlsaagpltfsmn 314
:
332 IANN-----KCAIAYIDGTSNKSISADRAHAIIFENEIVNTVNANGTSNAMPRRNAITY 387
:
315 rcgnlaagkggaladadgsisltaangdtfligntl-----tsstapstrnaIyl 366
:
388 ASSSGEILLGAGSSONLIYDP IEVSNAGS-----VSFNREADGTGSVFVGATVNS 440
:
367 gssakltnltraagqslyfllfniasnttgaaedvltlingdpnspldygslvfsegl-k 425
:
441 ADHIQR--NIOTKRPAPLTYSNGFLCIEDHAQLVYNRTQTGGGVYSLCGNAVLCSTYKKA 498
:
426 adekaadnfstlikqpialasgtalagnveldvngtlftegscll-----mqpgt 477
:
499 GNSASNASITLTKHGILSSILKSGAEIPLLAVPEPTNSNNVTADTAATFSIDVKLSLI 558
:
478 klkadetaistklkvdsai-----egnksvsietdganktiltfs-plvfyg 524

Db 678 dfeenifsvafcgldfgkdlfiven-tshnylaelylqhraflgplmpsfsgltdmlk 736

Oy 770 EGFLLTKLVGYSYGDHNCHEFT-----QGENLTSGQTFRSQTMGAVFFDLPMK-PFG 823

Db 737 dipllnaqlsytukndmdrlytsypaagswlnsgale---lgsalalyipkeaplf 793

Oy 824 STHTLAPFLGALGYSSLSHFTVEGAVPRSFSTKPLINLVLPICVKGSPFNATQRPQA 883

Db 794 qgyf---pflkfqavysrqnfkesgaearafd-dgdlvncslpvglrlekisedek-nn 848

Oy 884 WTVELAVQPVLYRQEPGATQOLLASKGIW 912

Db 849 feislavgydvyrknprsrtslmvsgasw 877

RESULT 23

W88424

ID W88424 standard; Protein; 930 AA.

AC W88424;

XX 26-APR-1999 (first entry)

DT

XX Chlamydia pneumoniae surface exposed protein Omp11.

DE

XX Omp11; outer membrane protein 11; surface exposed protein;

KM antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.

XX Chlamydia pneumoniae.

OS

XX MO9858953-A2.

PN

XX 30-DEC-1998.

PD

XX 19-JUN-1998: 98WO-DK00266.

PF

XX 23-JUN-1997: 97DK-0000744.

PR

XX (BIRK/) BIRKELUND S.

PA (CHRL/) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;

PI Mygind P;

XX WPI: 1999-105610/09.

DR N-PSDB: X06823.

XX

XX Species-specific test for identifying mammals infected with

PT Chlamydia pneumoniae - comprises detecting antibodies specific for

PT outer membrane proteins of C. pneumoniae or nucleic acids encoding

PT these proteins

XX

PS Claim 7; Page 63-65; 115pp: English.

XX

XX This polypeptide comprises the novel 97.6 kDa surface exposed

CC protein Omp11 of the human respiratory pathogen Chlamydia

CC pneumoniae. Its amino acid sequence was deduced from DNA (see

CC X06823) isolated from a C. pneumoniae expression library. The

CC invention provides 12 novel surface exposed proteins. Omp4-Omp15

CC (see W88417-28), and nucleic acid sequences encoding them (see

CC X06816-27). A new species specific test is claimed that is used

CC to identify mammals (including humans) infected with Chlamydia

CC pneumoniae. The test comprises detecting antibodies specific for

CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer

CC membrane proteins, especially by PCR. The proteins are also used

CC in the diagnosis of C. pneumoniae infection in mammals. The

CC nucleic acids and proteins can also be used in the immunization of

CC mammals, the nucleic acids being particularly useful as DNA

CC vaccines for effecting in vivo expression of antigens. The

CC vaccines may also prevent atherosclerosis and bronchial asthma,

CC which are possibly associated with C. pneumoniae.

XX

XX Sequence 930 AA;

XX

Query Match 12.0%; Score 608; DB 20; Length 930;

Best Local Similarity 26.9%; Pred. No. 1.4e-36;

Matches 250; Conservative 140; Mismatches 375; Indels 164; Gaps 40;

Oy 47 TGDTHNLNLCYIDNRLRYLALIQKTPNCGAAT---INDYLSFPTQKQEGIFYAKNLPE 103

Db 50 tada-nglnvylsgnvyi-----ndagkytalgccfletgdlffgkyfsfth-lvd 102

Oy 104 SCGAIGVASPNSPIVEIRDT---IGPIFENNCCRPFTSSNPAAANKIREGAIHAQ 159

Db 103 agnagaasttadkajlftgfnlsfiaapgtvaasgkstssagalnlnclngtllfsq 162

Oy 160 NLYINHNHDVGFMKNFYVRCGAISTANTFVSENOCSFLMDNICIOTNTAGK-GGAI 218

Db 163 nvsneannh-----galtt-kltslsgnts-----slftnsaakklggal 203

Oy 219 YAGTNSFESNCDLFTINN-ACCAGGAI-FSPICSLGNGNIVFYNNRCFKNVETASS 276

Db 204 yssaaaslsngtqglvlnmkgetlvggalgleaassltqn-sllfsgn-----lald 255

Oy 277 EASDGAIKV-----TTRLDVTGNRGRIFFSDNIRKNKGAIVAPVTLVNGPTPIFN 331

Db 256 aagkgaylcyektgetpeltlsgnks-ltfaensvtggaicahglolsaagpdlism 314

Oy 332 IANN---KGAIIYIDGTSNSKISADRHAIIFENENIVTNVNTANGSTTSANPPRRNATV 387

Db 315 rcgntlaagkggaladsgslsangsdlcflgnltl-----tstapltstnaly 366

Oy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAQVS-----VSFNKEADQGSVYFSQATVNS 440

Db 367 gssaklnltraagqsllyfdpilasnlvgasdvllingpdsnspldysglvlsqekl-s 425

Oy 441 ADFIQR--NLQTKTPAPVTLTNGFLCIEDHMQLYNRFOTRGVGVLSNGAVLSQYKKA 498

Db 426 adeekaadnftslkqplalasgtlalnkvneidvngftlqegslll-----mqpqt 477

Oy 499 GNSASNASITLKHIGLNLISLKSAGAEIPLLWVEPTNNSNNYADTAFTSLSDVKLSLI 558

Db 478 klkadleaalslklivdlasal-----egnkavslataganklltts-plvfq 524

Oy 559 DDYGNPVESTDLTHALSSQPMLSISEASDQNLRSDDMDPFGCLAV-----PHYGQGL 611

Db 525 dssgnf-yeshltlqatf-rqplivflaat-----aasdiyalallspvqrpehygy9h 578

Oy 612 WTGWAKTQDPEPASATITPOKANRFRHTLLTLWLPAGVPSPKHRSPLIANTLQNM 671

Db 579 weatwa---dstaksgt-----mtwvltgynpmpetrasvpsdlwaf 620

Oy 672 LLATESLKNASALETPSDHPFV---GITGGGLGMVYODPRENHPGFHRRSGYSAGMIAG 728

Db 621 ---cdirltqlqmtsqaanslyqfqlwasgylanffhkkksqlnqaftrkxsgylvgvsae 677

Oy 729 --QHTFSLKRSQTYTKINEXYAKNNNSKNVSCO-----GEMLSLQ 769

Db 678 dfeenifsvafcgldfgkdlfiven-tshnylaelylqhraflgplmpsfsgltdmlk 736

Oy 770 EGFLLTKLVGYSYGDHNCHEFT-----QGENLTSGQTFRSQTMGAVFFDLPMK-PFG 823

Db 737 dipllnaqlsytukndmdrlytsypaagswlnsgale---lgsalalyipkeaplf 793

Oy 824 STHTLAPFLGALGYSSLSHFTVEGAVPRSFSTKPLINLVLPICVKGSPFNATQRPQA 883

Db 794 qgyf---pflkfqavysrqnfkesgaearafd-dgdlvncslpvglrlekisedek-nn 848

Oy 884 WTVELAVQPVLYRQEPGATQOLLASKGIW 912

Db 849 feislavgydvyrknprsrtslmvsgasw 877

RESULT 24

W88417

XX WO200032794-A2.
 XX 08-JUN-2000.
 PD
 XX 01-DEC-1999: 99WO-CA01147.
 PF
 XX 01-DEC-1998: 98US-0110339.
 PR 01-DEC-1998: 98US-0110340.
 PR 01-DEC-1998: 98US-0110427.
 PR 01-DEC-1998: 98US-0110428.
 PR 01-DEC-1998: 98US-0110438.
 XX
 XX (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J;
 XX
 DR WPT: 2000-412339/35.
 DR N-PSDB: A30847, A30848.
 XX
 XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
 PT preventing, diagnosing and treating diseases such as community acquired
 PT pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
 PT asthma -
 PT
 XX
 XX Claim 16; Fig 1; 174pp; English.
 PS
 XX
 CC This sequence is a Chlamydia antigen of the invention, designated
 CC CPN100634. The nucleic acids (and their complementary sequences) may be
 CC used as diagnostic agents for detecting the presence of nucleic acids
 CC encoding Chlamydia antigens in samples according to standard methods,
 CC and therefore, for diagnosing Chlamydia infections. For example, they may
 CC be used as primers and probes for diagnostic polymerase chain reaction
 CC (PCR) assays. Antisense sequences may be used to down regulate
 CC expression of the proteins and may be used to treat infections. The
 CC nucleic acids may also be used to produce the protein antigens they
 CC encode according to standard recombinant DNA methodologies. The
 CC proteins may then be used as antigens for the production of antibodies
 CC (i.e. as vaccines) for preventing infection by Chlamydia. The
 CC antibodies may also be used as diagnostic reagents for detecting
 CC infections. Chlamydia is a pathogen implicated in the development of
 CC (for example) community acquired pneumonia, upper respiratory tract
 CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
 CC adult-onset asthma and acute exacerbations of asthma in adults.
 CC
 XX
 XX Sequence 928 AA:

Query Match 11.7%; Score 597; DB 21; Length 928;

Best Local Similarity 26.1%; Pred. No. 9, 2e-36;
 Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

QY 36 TKESLSNKISLTGD-----TNLITNCY--LDNLRVY-----LAILQKTPNEGA 76
 DB 45 lpktsatysltgdyfeyepgktpdsdcfktcdnlftlgnhsltfgfddaglhaga 104
 QY 77 AVTIDYLSFDTQKEGIFFAKNLTPESGCAIGVASPSNPTVEIRDITGPIVFENNCTCR 136
 DB 105 aastc-----anklftsgfsllstfsspsstv-----ltgq----- 136
 QY 137 PFTSSNPAAVNRKIREGAIHQNLYNNHNDVGFNMFNFSYVRGCAISTANTFVASENQ 196
 DB 137 -----gltsagagvlenl---rklyvag---nftsaadgaikga-sllltgts 178
 QY 197 SCFLPMDNICIOTNTAGGAIYAGTSNFSFENNCDL--FTINNACGAGAIISPICSLT 254
 DB 179 gdalifsm-----ssstkgal-altagariantgyvrflsnlastsgalddegtsil 232
 QY 255 CNRGNIVFYNNRCFNVETASSEASDCAIKVTRID---VTGNRGRIFSDNTITKNG 310
 DB 233 sonkfllyfegn-----aaktggaicntkassgsPELLISnktllfasnvaetsg 282
 QY 311 GAIVAPVTVLDNGPPTYFI-NNI--ANNKGAIYIDGTSNSKISADRHAIIFENIVTVNV 367

DB 283 gainhakkllalsggyfcellinnvssatpkygalsidaagelsisaetgnltfvrnlt-- 340
 QY 368 TNANGTSTANPPRRNAITVAVSSGCEILLAGSSOMLIFYDPI-EVSNAGVSVSFNKEA 425
 DB 341 -----ltgstddpkrnalngngkftelraaknhliffydphtsestsdvlikingaa 395
 QY 426 ----DOTGSVVFSGATVNSADFH-QRNLOTKTPAPLITLSNGFLCIEDHQAOLYVNRTO-T 479
 DB 396 galnpyggtlltsgetlidelkvadnlkssftqpvslsgklllqkyvtlstsfsgea 455
 QY 480 GGVSLSGNGAVLSCYKNGAGNSASNSITLKNHGLKSSILSGATPIPLMWEPTNNSNN 539
 DB 456 galldmgsdltl-----stagsliltnginvald---glkqv-----s 493
 QY 540 YTAADTAATFSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASONLRSDDMDS 599
 DB 494 ltkgasnkvivsgklnlidlagnl-yes----hmfsdqllslklvldadvtlnvds 548
 QY 600 GL-NVP-----HYGQGLWTGMAKTQDPERASSATITDPOKANRPHRTLLTLWPAGY 652
 DB 549 sllypaeqpnseygftqgwnvnm-----tlolatnkteat-----atwktgft 592
 QY 653 VPSPKHRSPLIANTLWGNMLLATESLKNSAEL--TPSDHP--FWGITGGGLMNYQDPR 708
 DB 593 vpspekksalvncnlwgv-ftdlrslqglveigatcmehkgfiv---vsmtnflnklgq 648
 QY 709 EHNHPCFMRRSGVSAQMIAGOTHT-----FSLKFSQTYTKLMERY-AKNN----- 752
 DB 649 entrgftrhsygy---vlgssahtrpkddlftafchlfardcdcfiahnnstlyggtlff 705
 QY 753 -----VSSKNYSCQGMELFSLQ--EGF-----LTKLVLYSYGDHNCHEFTYQGENLTS 800
 DB 706 khshtlqpnylirlgrakfsesaiekfpreipjaldivgsfshsdmrmethylsipe--s 763
 QY 801 OGFRFSRQGTGAVFDPDPRKPRFGSTHILTAFLPLGALG---IYSSLHFTFVAGAPRSFST 857
 DB 764 egswsneciaagldjldlpt-vlsnphlilktilipqmkvemvvsqnsfsssdgqfisl 822
 QY 858 KPIPLINLVPIGVKGSFNNATQRPQAMTVELA--YQPVLYROBPCGIATQLLASKGIWPCS 915
 DB 823 gr-llnlsipgak--flvg-dlgdsytyldsgffvysdvyrnmpgstallvmspdxwklr 878
 QY 916 GSPSSRHA 923
 DB 879 ggnlsrqa 886

Search completed: May 6, 2001, 19:16:43
 Job time: 4308 sec


```

Db 555 NPODMONTGTGTGTPSSSTSSISTPMIIFNGRLSIYDENVESYDSMDLSRGKADQLIIST 614
OY 584 SEASDNOLRSO---DMDFSGNLVPHYGWGLMTGMWAKT-----ODP PASSATITD 632
Db 615 ETTDGDOLDSSMOSSLNLSLSPHYGTGLMTPTMTTITTLNNSSAPTSATISIAE 674
OY 633 POKANR-----FHRLLLTWLPAGYVPSFKH 658
Db 675 OKKTSETFTPSNTTASIPNIKASAGSGSASNGEYITIKHTLVNMAPYGVLDPIR 734
OY 659 RSLPIANTL---WGMMLLATESLKSASALTPSDHPNCTGGGLGMVYQDPREHHPFH 715
Db 735 RGDLLANSLVHSGRMWTGLRSL-----PDNSWFALOGATTLFTKQQRRLSYHSTIS 787
OY 716 MRSSYSAGMIA--GQTFHSLKFSQTYTKLNERYAKNNVSSKNY---SCQGEMLFSIQE 770
Db 788 SASKGYTVSSQASGAHGHKFLSFSQSDKMEKETNNRLSRYYLSALCFHPWFD--- 844
OY 771 GFLTKLVG--LYSYGDHCHHFYTOGENTLSQTFRSQTMGAVFPL---PMKPFQS 824
Db 845 ---RALIGAAACNGYTHMRSFY--GTKKSSKGKHFSTTGLASLRCELRDMPRLR--- 895
OY 825 THILTAPELGALGIYSSHSFTEVGAYPPRSSTKTPILNVLVPIGVKGSFMATQRPQAW 884
Db 896 -SIMLTPFOALFSTRTEPASIRESGDLARLFTLEQAHTRAVSPIGIKGAYSSDTPITLSW 954
OY 885 TWELAYQVPLVNOEGDIAQMLLASKGFMFGSPSSRSMASYKISQOQPLMLTLFHOY 944
Db 955 EMBL:Y00544: BAA16052.1: ALT_INIT.
OY 945 HGFYSSTFCNLYNGEIALRF 965
Db 1014 QAEVATSTVSHYINMGALVF 1034

RESULT 3
YFAL_ECOLI STANDARD: PRT: 1250 AA.
AC P45508: P45507: P45506: P39441: P76468: P77487:
DT 01-FEB-1995 (Rel. 31, 'Created)
DT 01-NOV-1997 (Rel. 35, 'Last sequence update)
DT 01-OCT-2000 (Rel. 40, 'Last annotation update)
DE HYPOTHETICAL 131.2 KDA PROTEIN IN UBIS-NRDA INTERGENIC REGION
DE PRECURSOR.
GN YFAL.
OS Escherichia coli.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562:
RN [1]
RP STRAIN-K12 / MG1655:
RC MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12:
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Iemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Tagami H., Takekoshi K., Wada C., Yamamoto Y., Yano M.;
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
RX MEDLINE-84272624; PubMed-6087316;
RA Carlson J., Fuchs J.A., Messing J.;

```

```

RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
RT reductase operon."
RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
RN [4]
RP SEQUENCE OF 925-1198 FROM N.A.
RC STRAIN-K12 / EMG2:
RA Escip P., O'Keefe T., Robison K., Church G.M.;
RN Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE OF 1180-1250 FROM N.A.
RC STRAIN-OV6:
RX MEDLINE-88201664; PubMed-2834621;
RA Hussain K., Elliott E.J., Salmond G.P.C.;
RT "The parD-mutant of Escherichia coli also carries a gyrA mutation.
RT Mol. Microbiol. 1:259-273(1987).
RN [6]
RP IDENTIFICATION.
RX MEDLINE-96032851; PubMed-7567469;
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
RA Danchin A.;
RT "Detection of new genes in a bacterial genome using Markov models for
RT three gene classes."
RL Nucleic Acids Res. 23:3554-3562(1995).
CC CC -I- SIMILARITY: TO E.COLI YDEK.
CC CC -I- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
CC CC MANY FRAMESHIFTS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: AE000313: AAC75293.1:
DR EMBL: D90855: BAA16052.1: ALT_INIT.
DR EMBL: D90854: BAA16050.1: ALT_INIT.
DR EMBL: K02672: ? NOT_ANNOTATED_CDS.
DR EMBL: Y00544: AAA74094.1:
DR EMBL: Y00544: ? NOT_ANNOTATED_CDS.
DR Ecogene: E012850: yfal.
KW Hypothetical protein; Repeat: Signal.
FT SIGNAL 1 23
FT CHAIN 24 1250
FT DOMAIN 919 948
FT CONFLICT 28 30
FT CONFLICT 40 40
FT CONFLICT 65 66
FT CONFLICT 431 431
FT CONFLICT 433 434
FT CONFLICT 478 478
FT CONFLICT 773 773
FT CONFLICT 853 853
FT CONFLICT 923 924
FT CONFLICT 948 994
SO SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;

Query Match 4.3%; Score 220.5; DB 1: Length 1250;
Best Local Similarity 22.7%; Pred. No. 5,9e-06;
Matches 143; Conservative 84; Mismatches 235; Indels 167; Gaps 26;

OY 57 YLDNRITLAILOKTPREGAAYT-----ITDYLSPFPOKEGIV 95
Db 10 YLSLSPMISL-FSANGVAAYVDSGCGYVKASQASRSLSGITODMSIAGQ--WLV 66
OY 96 FAKNLTPESGCAIGYASPSNPSTVEIRDTICPVIFENNCTCRPTSSNPNAVKKIREGCA 155
Db 67 FSDMTNNAAGVAFLGQGAFLPENETGMTLFFANN-----VTGEYNNGA 114

```

```

Db 807 SADNTGTLFVNTPDI---TVTLNKGAGYFGVLYKOVIIISGPNIVFENICNGVIGHIA 863
Oy 248 SPICSLTGNRCNIVFYNNRCFKNVETASSEASDC-----GAIKVTRLDVYGNRGRI-- 299
Db 864 NSISEMNSLGTSLFLPGSTPLDVLTIKSTYNGCNVDNFNAPIVYVSGIDSMINNGQILIG 923
Oy 300 -----FSDN-ITKNNYGAIYAPVTVLYDNGPTFYFINNIANNKGAIYIDGTSMS- 348
Db 924 DKNIIALSLGSDNSITVN-ANTLYSGIRTKKNNOGTVTLSCGMNPNNGTIIYGLGLENGS 982
Oy 349 -KISADRRAIIFNE--NIVTVNTNAN-----GTSANPPRNAITVASSSGE- 393
Db 983 PKLKOVFTTDDYNLNGLSITANNVTINDVYTLTGGIAGDEDAK-----ITLGVNGNA 1036
Oy 394 -----ILGAGSSQNLIFYDPIEVSNA-----VSSEFKKADDTGSVY 432
Db 1037 NVRFVDSSTFSDPRSMIVATQAKKTVTLGNALVSNIGSLDTPVASVRF--GNDSCAGL 1094
Oy 433 FSGATVNSADPHQRNLQTKTPAPLTLSNGFLCIEDHAQTLVRFYQTGGVYSLGNAVLS 492
Db 1095 OGNIYSQNIIDFCTYNL-TLINSNVILGGGTTAINGEIDLTTNNLIIFANGTSTWCDNTSIS 1153
Oy 493 CYKNAG-----NSASNASITLK-----HIGLNLSSILKSGAEIPLLM 530
Db 1154 TTLNVSSGNIGOVVIAEDQAVNATTTGTTIKIODNANANESGTOAYTLIGGARF---- 1209
Oy 531 VEPTNNSNNYATDAATFSLSDVKLSLIDD-----VSGNSP 565
Db 1210 -----NGTIGARNFATVGTGNIIFVKYELLRDSNODVYLTFTNDVNLVYTTAVGNSAIANAP 1264
Oy 566 YESTDLTHALSSQPMKLSIASDNDLRSDMDPSGLNPHYGMOGLWYGMAKTODPEPA 625
Db 1265 GYSONISRCLES---TNTAAVNNMLAKDPDVATFV-----GAIAITDTSAAV 1309
Oy 626 SSATITDDPKARFHTLLTTLMLPAGYVSPKHSPLANTLGMNLLATFSLKNSAEIT 685
Db 1310 TVYNLMDTKOD-----LLSNRLGTLRLLSNAETSDVAGS---ATGAVSSDEDE 1357
Oy 686 PS---DHPFMCITGGCLGMVYVODPRENHPCFHMRSQSVYAGM--IAGOTHTFSLKFSQ 739
Db 1358 VSYGVAKKAFYNN-----AEQDKKGIAGAKKATTVGVVGLDITADSNLMTGAIGI 1409
Oy 740 TYTKL-NEYRANKNVS-----KNVSCQEMLFSLDEGLTKLVGLYSY 783
Db 1410 TKTDIKHODYKKDKDTDINGLSFSLYSQQLVKNFPAQCNALFTLTK-----V 1457
Oy 784 GDHNCHEFTTGCENLTSQ---GTFRSOTMGANFF--DLPMKFGSTHILTAFLCALGI 838
Db 1458 KSKSQRYFFESNKGMSKQIAAGNYDMNMTFGCNLIIFYDYINAMP-----NVLVTPMAGLSYL 1513
Oy 839 YSLSHTEVGV---AYRPSFTKPLINLVLPYGVKSGSEFNAT 878
Db 1514 KSSNENYKKEGTIVANKRINSKFSRVDLIVAKKAVAGSTVNIIT 1556

```

```

RC STRAIN=MLMINGTON;
RX MEDLINE=94040787; PubMed=8224886;
RA Hahn M.-J., Kim K.-K., Kim I., Chang W.-H.;
RT "Cloning and sequence analysis of the gene encoding the crystalline
RL surface layer protein of Rickettsia typhi.";
RL Gene 133:129-133(1993).
RN [2]
RP PARTIAL SEQUENCE.
RC STRAIN=MLMINGTON;
RX MEDLINE=92114896; PubMed=1370573;
RA Ching W.M., Carl M., Dasch G.A.;
RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii.";
RL Mol. Immunol. 29:95-105(1992).
RN [3]
RP IDENTIFICATION OF CLEAVAGE SITE.
RX MEDLINE=92104668; PubMed=1729180;
RA Hackstadt T., Messer R., Cleplak W., Peacock M.G.;
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent
RT mutant deficient in processing.";
RL Infect. Immun. 60:159-165(1992).
CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC S-LAYER WITH HEXAGONAL SYMMETRY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC EMBL: L04661; AAB48987.1;
CC DR Antigen; S-layer; Transmembrane; Cell wall.
CC KW CHAIN 1 1353 120 KDA SURFACE-EXPOSED PROTEIN.
CC FT CHAIN 1354 1645 32 KDA BETA PEPTIDE.
CC FT TRANSMEM 1415 1429 MEMBRANE ANCHOR (POTENTIAL).
CC FT CONFLICT 657 657 H -> N (IN REF. 2).
CC FT CONFLICT 842 842 V -> I (IN REF. 2).
CC FT CONFLICT 1071 1071 G -> A (IN REF. 2).
CC FT CONFLICT 1306 1306 G -> S (IN REF. 2).
CC SO SEQUENCE 1645 AA; 169698 MW; 0CB5641C7EB185EE CRC64;

Query Match 3.8%; Score 191.5; DB J; Length 1645;
Best local Similarity 19.9%; Pred. No. 0.00053;
Matches 206; Conservative 124; Mismatches 431; Indels 275; Gaps 44;

```

```

Oy 8 FLIGNSLSGL-AREVPSRFLFMPNSVPPDPKESLSKISITGTHNLNLCYDNLRYILA 66
Db 629 FNVSSSKTILNAGDVAINELWMEND-----GSVHLTHNTYL----- 664
Oy 67 ILQKTL---PREGAAVITTDVLSFDQKEGIVFAKNLTPESGGAIGVAPNSPTVEIRPT 123
Db 665 -IRKTNANOGKIITYAADPINDTALDQTNLSAESPLSNHFKTKANGSIT-LHIG 722
Oy 124 IGPVIFENNTCCRPFTSSNNAVNRKIREGAI-----HAQNLVYNH---NHQVGF 172
Db 723 KGVNLVANN-----ITTDANVGLSLHFRSGTISIVGTGCGOGLKLNLLIDNGTIVKF 777
Oy 173 MKNFSVYRGAIISTANTFVSEN-----QSC-----FLFMDNICIQNTAGKGAIV 219
Db 778 LQDITFGGKRIKIGKSLIOJSSNYITDHTIESADNTGTLFVNTPDI---TVTLNKGAGYF 834
Oy 220 AGTSNSFESNNCDLFF--INNACCAGAGIAPSPICSLTGNRGNIIVFYNNRCFKNVETASSE 277
Db 835 GVLKQYVNSGPGNIAFRNEIGNSVGAHAIAVDS-ISFENASLGASLFLSLGTPDLVLTIKST 893

```

```

OY 256 NRGNIY-----FYNNRCFKNVETAS-----SEASDGAIAVTRTLDVTGNCRGRI--- 299
DB 532 DAGNTMPLTIKSTVGNKTKAKGDFVSVVLGVDSVIADG-----QVIDONNIIVGL 582
OY 300 -FFSDNITKNVGAIVAPVTVLVNDGPTFYFINNIANKKGAIYIDGTSNKSISADRHAI 358
DB 583 GLGSDNGIIVNATTLVAGISTLNNNOGTVTLGSGVPTPGTVGLGTG---IGASK----- 635
OY 359 FENIVTNTVANGSTSNAPPRNAITVASS-----SGEILGA-GSSONLIFYPDI 410
DB 636 FKQVFTTIDYNNLGNIIATNATINDGVYTTGGIAGIGDGIITLGSVNGNNGVRFADGI 695
OY 411 EVSNAGVSVSFNKADOT-----GSVFSGATVNSA----- 441
DB 696 -LSNSTSMIGTKRANKGTIVYLGNAFVGNIGSDPFVASFRTGSDSGAGLOGNIYSQVI 754
OY 442 DEHORNLOTKTAPRLTSLNGFLIEDHQAOLVNRFTQTGCVSLGNG---AVLSCTYKNG 497
DB 755 DFGTYNLGI-VNSNIIIGGTTAINGKIDLVNTLTFASGTSTWGNNTSIEFTTLANGN 813
OY 498 AG-----NSASNASITLK-----HIGLNLSSILKSGAEI-----PLLWEP 533
DB 814 ICHIIYILECAQVNTTTTGTITTIKVODNANANFSGTOTYTLIOGARFNGLTSPNFAVYG 873
OY 534 TNNSNVYTAADTAAT-----FSLSDVYKLSLIDYGNSPY-----ESTDLTHAISOPM 580
DB 874 SNRFVYVSLIRANQDYVITRTNMAENVVTTNDIANSFPGAGPQVDQNTTEVNATTAAY 933
OY 581 LSISEASDNLQSDDMDFSGLVNPHYGWGLTWGAKTQDEPPASATITDPO----- 634
DB 934 NNLLAKNS-----ANSANFVGAIVTD-----TSAITVYQDLDLAND 970
OY 635 -FANRHFLLLTWLPAGVVPSPKHSPLIANTMGNNMLLATESKNSAELTFPSDHPFG 693
DB 971 IQAQQLNRLGALRYLG---TPETAEMAGPEACAIANAANGDAIDONVA-----YG 1018
OY 694 ITGGGLGMVYODPRENHQGFHMRSSGYSAGM---IAGQTHFSFLKFSQTYKL-NERYAK 750
DB 1019 IMAKPPTYDAHOGKKGGLAGYKAKTGTGVYIGDLTDLANDMLCAIGITRTDKHODYK 1078
OY 751 NNVS-----KNYSCGEMLFSLQEGFLLTKLVGLXYGDHNCNHHIYTOG 795
DB 1079 GDKTDVNGFSFSLYGAQOLVKKNFAQGSALFSLNQ-----YKKRSQRYFFDAN 1126
OY 796 ENLTSG---CTFSQTMG---AVFEDLPKMPFQSTHILAPFALGIYSLSHFEVGA 850
DB 1127 GNMKSQIAGHYDMTFGGLVIGIDYN---AMOGVLYTPMAGLSYLKSDENYKETGT 1182
OY 851 -----YPRSFSTKTPLLINLVPIGVGSFMATQRPQAMTVELAYQPVLYR----- 896
DB 1183 TVANKQVNSKFSRDTDI---VGAKVAGSTMNT-----DLAVYEVHAIVHVKVTG 1231
OY 897 -----OEGCITQLLASKGIMFGSGSPSRHAKSKISQOQPLSMLT 939
DB 1232 RLKSTQSVLDGOVTPCINQDRTTKTSYNLGL---SASIRSDAKMEGXIGYDAQISSKYT 1288
OY 940 LH 941
DB 1289 AH 1290

```

RESULT 7

```

OMPB_RICRI STANDARD: PRT: 1654 AA.
AC Q53047:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE).
GN OMPB.

```

```

O5 Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=783;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-R;
RX MEDLINE=9216/7802; PubMed=1724278;
RA Gilmore R.D. Jr., Clepik W. Jr., Policastro P.F., Hackstadt T.;
RT "The 120 kilodalton outer membrane protein (Romp B) of Rickettsia
RT rickettsii is encoded by an unusually long open reading frame:
RT evidence for protein processing from a large precursor."
RL Mol. Microbiol. 5:2361-2370(1991).
RN (2)
RP SEQUENCE OF 279-1654 FROM N.A.
RC STRAIN-R;
RX MEDLINE=90136087; PubMed=2515418;
RA Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT "Cloning, expression and sequence analysis of the gene encoding the
RT 120 kD surface-exposed protein of Rickettsia rickettsii."
RL Mol. Microbiol. 3:1579-1586(1989).
CC - FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC - SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X16353; CAA34403.1;
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1333 120 KDA SURFACE-EXPOSED PROTEIN.
FT DOMAIN 1334 1654 32 KDA BETA PEPTIDE.
FT DOMAIN 1181 1188 POLY-THR.
SO SEQUENCE 1654 AA; 168184 MW; D7AB70FB7087F618 CRC64;

```

```

Query Match 3.7%; Score 190; DB 1; Length 1654;
Best Local Similarity 19.4%; Pred. No. 0.00066;
Matches 198; Conservative 130; Mismatches 424; Indels 270; Gaps 42;

```

```

OY 92 EGI-YFAKNLTPESGCAIGASNSPTVEIRDTIG-----PIVFENNTCCR-----P 137
DB 721 EGVNLVATNTTTTIDANVGSVFNAGGTNTVSGTVGQCGKRFVTALENQTYKFLGNAT 780
OY 138 FTSSNPAAVANKIREGGAIAHONLYINHNDVGFPMKNEFVYVGAISTANFVSENO 197
DB 781 FNGNTTIAANSTLQIG-----NTADCYA-----SADGTGIVEFNTPETVTLN 826
OY 198 CPLEMDNICIQITNACKGAIV---ACTSNSFESNCDLFFINNAACGAGIISPJCSLTG 255
DB 827 KOAPVVALKQITVSGPGNVVINEIDGNAGNHGAVDTITAFENS-SLGAVVFLPRGIPFN 885
OY 256 NRGNIY-----FYNNRCFKNVETAS-----SEASDGAIAVTRTLDVTGNCRGRI--- 299
DB 886 DAGNTMPLTIKSTVGNKTKAKGDFVSVVLGVDSVIADG-----QVIDONNIIVGL 936
OY 300 -FFSDNITKNVGAIVAPVTVLVNDGPTFYFINNIANKKGAIYIDGTSNKSISADRHAI 358
DB 937 GLGSDNGIIVNATTLVAGISTLNNNOGTVTLGSGVPTPGTVGLGTG---IGASK----- 989
OY 359 FENIVTNTVANGSTSNAPPRNAITVASS-----SGEILGA-GSSONLIFYPDI 410
DB 990 FKQVFTTIDYNNLGNIIATNATINDGVYTTGGIAGIGDGIITLGSVNGNNGVRFADGI 1049
OY 411 EVSNAGVSVSFNKADOT-----GSVFSGATVNSA----- 441

```


01-NOV-1997 (Rel. 35; Last annotation update)
DE VACUOLATING CYTOTOXIN PRECURSOR.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OX Helicobacter.
RN NCB1_TaxID=210;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=95355366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vaca types with cytotoxin production and
RT peptic ulceration."
RL J. Biol. Chem. 270:17771-17777(1995).
CC - FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC - SUBCELLULAR LOCATION: SECRETED.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U29401; AAA66834.1; -
DR CYTOTOXIN; Toxin; Signal.
FT SIGNAL 1 30
FT CHAIN 31 2
FT PROPEP ? 1310
FT SEQUENCE 1310 AA: 141988 MW: 18C21FE3D435F981 CRC64;
Query Match 3.4%; Score 173; DB 1; Length 1310;
Best local similarity 21.2%; Pred. No. 0.0054;
Matches 225; Conservative 137; Mismatches 374; Indels 360; Gaps 60;
QY 7 FFLIGNSL-SGLAREVPSRIELMPNSVPDPTKSLSKISL-TGDTHTNLTN----- 55
DB 209 FVEINRVGSGAGRKASSTYLTKSEKITSRENA--EISLYGATLNLVSSNOSVDLY 266
QY 56 --CYLNLRLYLILLOKTPREGAAVITIDVLSFPDQK-EGIFAKNL-----TPESGA 107
DB 267 GKVMGRLQVGAIVLAPS-----YSTIDTSKVOGEMNFRHLAVGDONAAOAGI 314
QY 108 IGYASPSPTVEIERTDTIGPYI-----FENNTCRPFSTSSNPNAAYNKIR----- 151
DB 315 IANKKTNIGLIDMOSAGSLITPRPEGVSESKTKDNPNKMDAQTEIQPOVIDGPF 374
QY 152 EGGAIHAQNLV-INNHND---VVGPMKNFS-----YVGGALSTANTVSVSENOGCF 200
DB 375 AGGKDTVNVFHNTRKADGTLRAGGFKASLSTNAHLHIEGGVNLN-----QASGRTL 429
QY 201 FMONICIGTNTACK-----GGAIYAGTSNSPF-----SNNOCLPFTNNCAGCAIF 247
DB 430 LVEMLTGNTVEGTLRVNNNOVGGAALGSSANFEFKAGEDTNNATATF--NNDIHLKAVN 488
QY 248 SPICSLTGN-RGNI-----VFYNNRCFKNVTASSE----- 277
DB 489 LRVDAHTANFNNGNIYLSKSTNLNRNGHTAFKFNIDATKSONGLTSTLDFSGVTDKVNIN 548
QY 278 ---ASDGAIAK-----VTRLDVTGNRGRIFESDNI-TKNYGAIA-----YAP--- 316
DB 549 KLTAAATNVIKKNFDIKELVYTVTFVOSGQ--YTFIFENIGDKSRIGVSVLOTGYSPAYS 606
QY 317 -----VYTLVNGP-TYF-----IN-----NINNNKGAIYIDGTS 346
DB 607 GGVTFKGKKLVYDIEYHAPNFTDARNVTDVEINKRILFGAPGNIA-GKTGLMFMNLT 665

QY 347 NSKISADRAHAIFFENIYVNTVNTANCTSTANPPRNATVYSSSGEIT-LIGAGSSQNL 405
DB 666 NSNASMD-YGKDDLDTLIOGHFTNNOC-----MNLFEVODRVAITLNAHQASMI 713
QY 406 -----FYDPIEVSNAGVSVFKNKADDTGS-----VFESGATVNSADFHQRNL 448
DB 714 FNNLDSTGTGFKPLIKINNAQNLTKNKEHVLKANKINDINLVGVGASVDNISASNTNL 773
QY 449 QTKTPAPLTLNSG-----FLCI--EDHAQLTVNRFOTGCVSLGAVLS----- 492
DB 774 QEPFKRLALYNNNNMDTCVVRKON-----LNDIKACG--MAIGNQSVNPNENYKYLEG 827
QY 493 -CYKN-GAGNSASNASTITKHIGLNLSSLSKSGAELPLLVNPEPTNNNNYATDATTESL 550
DB 828 KAKNKGCIKTKANTTIAV-NLC-NNSTPTNSTDTTNL--PTNTNN-----ARFA- 875
QY 551 SDVKLSLIDYGNPY--ESTDLTHALSSQPMLSIASDONLRSDMDPFGSLNVPHYGM 608
DB 876 ---SYALIK--NAPFAHSATPVLVAIINHDEGTIESVELANRSSDIDTLVANSAGOR 929
QY 609 QGLMTW-----GWAKT-----QDEPPASSATITDPKANRFRHTLLTLPLAG-- 651
DB 930 DLQOTLIDSHDAGYARTMIDATSAEITQOLNAATTLNINIASLEBKTSGLQTLISNA 989
QY 652 -----YVSPKHSPLIANTLM 668
DB 990 MINSRLVNLRRKHTNHIDFAKRLQALDORPASLSAEVLYOPAPYERK--TNW 1046
QY 669 GNMILATESLKNSAELTPSDHPWGITGGLGMVYODPRENHPGHMSSSGACMIAG 728
DB 1047 ANAIGCT-SLNG-----SNASLYG-TSAGVDAYLNGEVAIYGGF--GSYGIS-- 1092
QY 729 QTFHTFLKSQTYTKLNERYAKNN--VSSKNYSQGEMLFSLQ-----EGFL 774
DB 1093 -----FSNQANSLNSGANNTNFGVYSRI-FANQHEFDEAGALGSDSLSNFKSALL 1144
QY 775 TKLVGLSYGDHNCHEFTYOGENLTSOGTFRSOTMGANVF--DLPMKP-----FGS 824
DB 1145 QDLNOSYHY-----LAYSATTIRASYGYDAFFRNALVLKPSVGVSYNHLGS 1190
QY 825 THILTAFLGALGIYSSLSHFEVGA 850
DB 1191 TNRKSNNOVALSNGSSOHLFNANA 1216
RESULT 12
ID YS89_CAEEL STANDARD; PRT: 2329 AA.
AC 009624;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN ZK945.9
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCB1_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilkinson-Sprout J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC

Query Match 3.2% Score 161; DB 1; Length 1286;
Best Local Similarity 20.6% Pred. No. 0.029;
Matches 219; Conservative 115; Mismatches 359; Indels 370; Gaps 61;

69 OKTRPEG-AAATITDTLSEFDTOKECITFAKMLTPESGAI-----GYASPSPIVEIRD 122
Db 349 OHIRNGIASGATVINGSVINISGG--YAEETIINSGLTLEVLSDYA-----RG 397
OY 123 TIGPVIFFENNTCCRFPTSSNPAAVN-KIREGAIHAOLYINHHNDVYGFKNFSYVR- 180
Db 368 TI-----LNSGRE--NNSNGCVSYNAMINTGC-----NOIYSOCEATAIVNTSGGR 445
OY 181 ---GGAISTANTFVVSSENOSEFLMDNICIQNTAGK--GGAIVAGTNS-----FES 228
Db 446 INSGTAPVQNSVVTFRVVS-----SAKPFDAEVYSGKQTYLWLGIMWIS 492
OY 229 NNCDLFFINNACCAGGATFSPICSLTGRNGNIV--FYNNRCKNVEPTASSEASDGAIV 286
Db 493 NFLTAVSMFPGTAGANVNLGRLNAFAGNVGTILNDEGQVYVSGATATSTVG--- 548
OY 287 TRLDVTGMRGRIF-----FSDNITKNYGG--AIYA---PVYTLVNDGPTFYI----- 329
Db 549 -----NNEGEVYLSGCTIDGTIVLNSGGLAVSSGGKASATVINEGGAQVYVYDGGQVT 601
OY 330 -NNIANNKGAIYID-GTSNSKISADRAHAIIF-----NENIYTNVTN 369
Db 602 GTNININ--GGTIRVDSCASALNIALISSGGLNLTSGATIPELTTMAALSVSQNHASNIYL 659
OY 370 ANG-----TSTANPPRR-----NAITYASSGCEILLGAG--SSONLIF- 406
Db 660 ENGGLRLVTSAGTADDTIVNSAGRLRIDGDTINGITTTINAGC-IVAGNINIONDGFILN 718
OY 407 ---D-PIEVSNAGVSFN-----KEAD---QTSVVSFGATVSADEHQN 447
Db 719 LAENYDFEELSESGVLYVNDNTGIMTYACTLYOAGCVNKNKGIIIFDSAVVN-ADM--- 773
OY 448 LOTKPAPLTSLNGELCIDHAQLTVNRTOTGVVSLNGAVLSCYKNGAGNSASNASI 507
Db 774 -----AVNQNAVYINISD--QATINGSVNNNGSIVINNSII-----NONTITADAL 816
OY 508 ---TLKHIGLNSLILSCAEIPLWVEPTNNS-----NNTADTATFSLISDY---K 554
Db 817 SEGTAKLISATVNSGLVNNKNIIL--NPTKESAGNTLVSVYTGTPGSVISLGGLEGD 873
OY 555 LSLIDY---GNSPESTDLTHALSQPMLSISEASDNLR-----DDDMDFSL 601
Db 874 NSLTDRLVVKNTSGOS-DIVYV-----NEDSGGOTROGINIISVEGNSDAEFSLK 924
OY 602 N---VPHYGM-----OGLMTWGNAKTOD-----PEPASATITDPOKANRFRH 641
Db 925 NNVAAGAYDTLOKNGESTDNKGLWLTSLPTSDTRQYRPEMGSAAT----- 972
OY 642 TLLLTWLPAGVPSRKHNSPLIANTLMGNMLLATESLKSASBLT--PSDHPFW-GITGG-- 697
Db 973 -----NMLANSLEFLMDLNRKOFRAKMSDNTOPESASVMMKTIIGIS 1014
OY 739 QTYTKLERNYAKNNVSSKNYSCOGEMLFSLQEGFLTLTKVLGLYSYSDHNCHEFTYGENL 798
Db 1067 YTSNKA-----ARITLD--GYS-----VCYVG-----TWYONGENA 1095
OY 799 TSGGTFRSQTMGAVFDDLPKPFGSTHILTLAPLGCALCIYSSLSHTEYGAIPNSFSTR 858
Db 1096 T--GLF-AETWQYMFNFASVSKDG---LEEEKYNLNLITASAG-----GGYNLWVHW 1143
OY 859 TPLINVLVIGVGSFMANNTORPOAWTVELAYOPLVYROBPJATOLLASKGIMFGSGSP 918
Db 1144 TS-----PEGITGEFW---LDPHLOAVMMGVTPTDHEDNGTVVQAGCKNNIQTACIGR 1194
OY 919 SS-----RHMSYKISQTOPLS 936

Db 1195 ASWKVSTLKDXTGRFRFPYIEANNIHTHEPGVXMSDQSOLS 1238

RESULT 14
PK2-DICDI
ID PK2-DICDI STANDARD: PRT: 1858 AA.
AC P54674;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PHOSPHATIDILINOSITOL 3-KINASE 2 (EC 2.7.1.157) (PI3-KINASE)
DE (PDIINS-3-KINASE) (PI3K).
GN PI3K OR PIK2.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96009592; PubMed=7565716;
RA Zhou K., Takegawa K., Emr S.D., Firtel R.A.:
RT "A phosphatidylinositol (PI) kinase gene family in Dictyostelium
RT discoideum: biological roles of putative mammalian p110 and yeast
RT Vps34p PI 3-kinase homologs during growth and development.";
RL Mol. Cell. Biol. 15:5645-5656(1995).
CC -I- CATALYTIC ACTIVITY: ATP + 1-PHOSPHATIDYL-ID-MYO-INOSITOL - ADP +
CC 1-PHOSPHATIDYL-ID-MYO-INOSITOL 3-PHOSPHATE
CC -I- SIMILARITY: BELONGS TO THE PI3/PI4-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U23477; AAA85722.1; -
DR DictyDB: DD01100; PI3K.
DR InterPro: IPR000341; -
DR InterPro: IPR000403; -
DR InterPro: IPR001263; -
DR InterPro: IPR002420; -
DR Pfam: PF00792; PI3K_C2; 1.
DR Pfam: PF00794; PI3K_Rbd; 1.
DR Pfam: PF00613; PI3K; 1.
DR Pfam: PF00454; PI3-PI4_Kinase; 1.
DR PROSITE: PS00915; PI3_4_KINASE_1; 1.
DR PROSITE: PS00916; PI3_4_KINASE_2; 1.
DR PROSITE: PS50290; PI3_4_KINASE_3; 1.
KM Transferase; Kinase; Multigene family.
FT DOMAIN 34
FT DOMAIN 166
FT DOMAIN 172
FT DOMAIN 185
FT DOMAIN 226
FT DOMAIN 227
FT DOMAIN 235
FT DOMAIN 246
FT DOMAIN 253
FT DOMAIN 259
FT DOMAIN 268
FT DOMAIN 294
FT DOMAIN 303
FT DOMAIN 361
FT DOMAIN 364
FT DOMAIN 369
FT DOMAIN 384
FT DOMAIN 425
FT DOMAIN 429
FT DOMAIN 439
FT DOMAIN 444
FT DOMAIN 445
FT DOMAIN 454
FT DOMAIN 562
FT DOMAIN 570
FT DOMAIN 715
FT DOMAIN 727
FT DOMAIN 982
FT DOMAIN 990
FT DOMAIN 1015
FT DOMAIN 1049
FT DOMAIN 1598
FT DOMAIN 1858
SO SEQUENCE 1858 AA; 203945 MW; A6C033304CDEA03 CRC64;

Query Match 3.1% Score 159.5; DB 1; Length 1858;

```

Db 452 IGVA-----GCVN-----LNTSDCKEYSON-----ITFLGKASGIGTL 486
OY 439 NSADFHQNLQTKTPAPLTLSNGFLCIEDHQAOLTVNRFOTGCVVSLGCAVLSCKYNG- 497
Db 487 NLMD-----ATSSF-----DYGVINGNF-----GSGIVNNGCATLNTGTGCF 525
OY 498 ACNSASNAITLKHIGLNLISLTKSGAEIPLIWEPTNNSN-----NYRAD 543
Db 526 IGNASNGKI-----VNIST-----DSLMLNKTSTNQLLOVGLTGEINLT-- 569
OY 544 TAATSLSDVKSLLD-----DYNSPYESTDLTHALSOPPLSTISEADNGLRSD 595
Db 570 TGGIVKARTQIALNDKSKGVDRVGDQNSLLETENMYVGTSGTGLTLTNNGTFLNVEGGE 629
OY 596 MDESGLANPHYG-----MOGLWTGWAKTQDEPPASS 627
Db 630 V-YLGVFEPVAVGTNLGAHGEAADAGTITNATVEFGLGCVFVFNHTNNSDAGQYVD 688
OY 628 ATITDPOKANRFHRTLTLTWLPAGYVVPSPKHSPLIANTLMGNMLAEKLSAEIIPS 687
Db 689 MLITGDDKDKYIHDAGHTVFNAG-----NTYSG-----KTLVNDGLITIA 729
OY 688 DHPFGITGGGIGMMVYDOP-----RENHGFH-----MRSSGVSAGMIAGQ 729
Db 730 SHTADGVYGMGSEVYTIAPGTLDILASTNSAGDYTLTNALKGDLIRYQULSSDKMFGF 789
OY 730 TTTFSLKFS-----QYTKLNERAKNNVSSKNVSCGEMLFSLQEGFLTKVLGLYS 782
Db 790 TIAATGEFAGVQLKDSTTT-----LERNTAALITHA-----MLQSEVTTSVK-VCEDS 839
OY 783 YGDHNCHEFTYTGENTSCGTFERSQTMGAVEF--DLPKPKFGSHIILAPPLGALCIYS 840
Db 840 ICGLAMN-----GCTIIFDDIPATLAEIGYSVDTLVVAGADYT 879
OY 841 SLSHFEVCAVYRSPSTKPLINLVPIGVKGSFMAATROPAVYELAYOP-----VLX 895
Db 880 WKGRNYOVN-----GTGDVLIDY-----PAFMDPMKANNPLTTLNLE 917
OY 896 ROEPGIATOLLASKGIWFGSGSPSRHANSYKISQOTQPLSMTLTFHYUG 946
Db 918 HDDSHVGVLVKAKQYVIGSGSLTLNLDGDEVEADK-----TTHIAONG 962

```

```

RESULT 16
YPIA_ECOLI STANDARD: PRT: 1569 AA.
AC P52143: P76610: P77017: P77019:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 98.4 KDA PROTEIN IN ALPA-CABD INTERGENIC REGION (F949).
GN YPIA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=K12:
RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
RA Ikemoto K., Inada T., Isono S., Itoh T., Kanai K.,
RA Kasai H., Kashimoto K., Kim S., Kimura S., Kitagawa M.,

```

```

RA Kitakawa M., Makino K., Masuda S., Miki T., Mizobuchi K., Mori H.,
RA Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Tagami H., Takemoto K., Wada C.,
RA Yamamoto Y., Yano M.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC - SUBCELLULAR LOCATION: OUTER MEMBRANE (POTENTIAL).
CC - SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U36840: AAA79815.1: ALT_SEQ.
CC EMBL: AE000350: AAC75695.1: -.
CC EMBL: D90889: BAA16514.1: ALT_INIT.
CC EMBL: D90890: BAA16518.1: ALT_INIT.
CC EMBL: E613213: YPIA.
CC Ecogene: E613213: YPIA.
CC Hypothetical protein: Outer membrane.
CC SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;

```

```

Query Match 3.0%; Score 155; DB 1; Length 1569;
Best Local Similarity 19.7%; Pred. No. 0.086;
Matches 194; Conservative 125; Mismatches 320; Indels 344; Gaps 51;

```

```

OY 97 AKNLPESCGALGYASPNPVEIRDTIGVIFENNTCCRPSSNPN- - - - -AVN 148
Db 579 AKHVEQSGCALASTSTGTLIEGTNSYGDATYIRNSEAKNVLENAGSLTYVTSRAVD 638
OY 149 KI-----REGAI-----HAONLYNNHND----- 168
Db 639 TIINANGKMDYGVKDVGYLNSAGTQTIYASATSDKANIKGQKQYVGLATEANIESGEO 698
OY 169 -VGFPMKNSYVRCALIS-----TANFVSENSCFLFMDNICIQTTAGKGAIVNG 221
Db 699 IYDGGSTEXTHINGTQTOYQNGKAINTDIVSGLQ-----IMAGTAEGLIING 749
OY 222 T-----SNSFESNCDLFFINNACCAG-----GAIFSP--TCSLTGNNGNIVFY- 263
Db 750 SOVYNEGGIAENSVLNDGGLDYREKGSATGIGQSSQGLAVATRTATRTGTRADVAAS 809
OY 264 -----NNRCFKN--VETASSEASD-----GGAIKVTRLDVGN-----RGR 299
Db 810 IEGGANNILLANGLVLTYESDSSDKTQVNMGGREIVTKATATGTLTGGEQIVEG-- 867
OY 300 FESDNTIKYNG-----AIYAPVYLVNNGPYFI-----NIANNKGAIIYDGTNS 348
Db 868 -VANETTINGDIOTVSANGAEAIKTKINEGGLTVNDNGKATDIYONSSAALDTSTANI 926
OY 349 KISADRAHIIFF--NENIVNVTVANGSTSANPRNATIVASSG--EILLAGSSONLI 405
Db 927 ELISGTHQYGTFSISGNLATMMLLENG-----GNLVLAETEARSDTVGGGMQNL 977
OY 406 FYDPIEVSNAGVSVSFNKDAQ-----TGSVFSG-----ATVNSADFHORN 447
Db 978 GQDSATKRVNSCGGYTLGRSDEFGALARAEDLQVAGCTAIVAGTALDASVSGA--TGS 1034
OY 448 LQTKTP---APLTLSNGFLCIEDHQAOLTVNRFOTGCVVSLGCAVLSCKYNG- 497
Db 1035 LSLMTPRDNTVPKL-EGAVRITDSATLT-----LQNGVDITLADLADSRGS 1081
OY 498 -----ACNSASNAITLKHIGLNLISLTKSGAEIPLIWEPTNNSNMYVADTATFSLSDV 553
Db 1082 VMLNSNNSCAGTS-----NCEYRVNSLNDGQVYL-----SAQTAAPATTNGI 1125
OY 554 KLSL-----IDYGNSPYESTDLTHALSOPPLS-----ISEADNOLR-- 592
Db 1126 YNLTTNELSGSNF-YLHTNVAGSRGDQLVYNNNNTGNKIFVQDTGVSPQSDAMTLV 1184

```

DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).
 GN SLGA.
 OS Methanothermobacterium sociabilis.
 CC Archaeae: Euryarchaeota: Methanobacteriales: Methanothermobacteriaceae;
 CC Methanothermobacter.
 CC NCBI_TaxID=2181;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC STRAIN=DSM 3496 / KFI-FL;
 CC MEDLINE=91293115. PubMed=1712296;
 CC Broeckl G., Behr M., Fabry S., Hensel R., Kaudewitz H., Biendl E.,
 CC Koenig H.;
 CC "Analysis and nucleotide sequence of the genes encoding the surface-
 CC layer glycoproteins of the hyperthermophilic methanogens
 CC Methanothermobacter fervidus and Methanothermobacter sociabilis.";
 CC Eur. J. Biochem. 199:147-152(1991).
 CC or send an email to license@lsb-sib.ch).
 CC
 CC EMBL: X58296; CAA41229.1;
 CC PIR: S16375; S16375.
 CC
 CC Glycoprotein: Cell wall; S-layer; Signal.
 CC SIGNAL 1 22
 CC FT CHAIN 23 593
 CC FT CARBOHYD 29 29
 CC FT CARBOHYD 58 58
 CC FT CARBOHYD 66 66
 CC FT CARBOHYD 74 74
 CC FT CARBOHYD 114 114
 CC FT CARBOHYD 122 122
 CC FT CARBOHYD 145 145
 CC FT CARBOHYD 148 148
 CC FT CARBOHYD 158 158
 CC FT CARBOHYD 176 176
 CC FT CARBOHYD 208 208
 CC FT CARBOHYD 231 231
 CC FT CARBOHYD 326 326
 CC FT CARBOHYD 336 336
 CC FT CARBOHYD 340 340
 CC FT CARBOHYD 431 431
 CC FT CARBOHYD 471 471
 CC FT CARBOHYD 500 500
 CC FT CARBOHYD 516 516
 CC FT CARBOHYD 593 AA: 65503 MW: DFCDFB357EL0FZEB CRC64:
 CC SEQUENCE

Query Match 3.0%; Score 153.5; DB 1; Length 593;
 Best Local Similarity 22.9%; Pred. No. 0.029;
 Matches 119; Conservative 46; Mismatches 189; Indels 165; Gaps 28;

DB 92 EGIYFAKNLTPESSGGA--IGYASPSNPTVEIRITIGPVIEENNTCCRPFTSSNPNNAVVK 149
 DB 61 ENLVNKSILTKSNGSAIINAASSEKSTIRK-----ANNVWIEGF----- 101
 DB 150 IREGG--AIHAONL---YINNHVDVGFEMKFSYVRGAI-----STANTF---VSENOG 197
 DB 102 IITGGAKGIMENTGCTITNNITQNAFVSGMEYGGNGICLVNSTNITNNIIRNTW 161
 DB 198 CPLEMDNIC-----IQTNTAGKGG--AIYAGTSNFSFENN-----CDLF- 224
 DB 162 NGI---NVCESKGNIIKNNTIMYSGIGIYVWGFKFEGNNIENNRINATYGGIYLF 218
 DB 235 -----FINNACCGAIAFSPIC-----SLTGNRGNIY 261
 DB 219 PSNNKICRNTIANVSSGGGMSGALCIDVSDYNIWKDNIGVNCDDGLFTDGMIGNE----- 274

QY 262 EYNNRCFKNVETASSEASDGAIKVTRLDVTGNRGRIFFSDNTITKNYGAIYAPVTV 321
 DB 275 -ITNNIFKCKVAVSESTYGA-----SRNNKIY--GNFIFYETAISDPKGLV 321
 QY 322 DN-----GPTFYFINNNNGCAIYIDGTSNRSKISADRHAIIF-----ENVT 365
 DB 322 DNIMWTBEGNYSWNTGNNGT-----DOTGNIPIYYDKPLVDLAIEDIAKPSGIEV 376
 QY 366 NTYNNANGTSTSANPPRNAIYVASSGCEL-----LGASSQNLIFYDPLEVSNAVS 419
 DB 377 RKNLCKADIKKIDPLTKIKISCDNDYERFIPLPSGEGQ-IVRWDKI-VPEGNHTI 434
 QY 420 SFNKADQTSVFSVSGAIVNSADPHQRNLQTPAPPLSLNGFELCIEDHAQLTVNRPTQT 479
 DB 435 KAEIPYSAGYLI--GTNIRDADI--SNVFSK-----MVQGF-----YONKTF 475
 QY 480 GGVSLGNGAV-----LSGYK---NGAGNSASNASTLK 510
 DB 476 ITLTLGKSTITIKYIISITVMPVNGTKVSYRELITTLK 514
 RESULT 19
 FTBP_ADEB3 STANDARD: PRT: 976 AA.
 ID FTBP_ADEB3
 AC 003553;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE FIBER PROTEIN.
 GN PIV.
 OS Bovine adenovirus type 3 (Mastadenovirus bos3).
 OS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 CC NCBI_TaxID=10510;
 CC (1)
 CC SEQUENCE FROM N.A.
 CC MEDLINE=93107871; PubMed=1469367;
 CC Mittal S.K., Prevenc L., Babluk L.A., Graham F.L.;
 CC "Sequence analysis of bovine adenovirus type 3 early region 3 and
 CC fibre protein genes.";
 CC J. Gen. Virol. 73:3295-3300(1992).
 CC [2]
 CC REVISIONS.
 CC MEDLINE=94103794; PubMed=8277294;
 CC Mittal S.K., Prevenc L., Babluk L.A., Graham F.L.;
 CC "Sequence analysis of bovine adenovirus type 3 early region 3 and
 CC fibre protein genes.";
 CC J. Gen. Virol. 74:2825-2825(1993).
 CC [3]
 CC SEQUENCE FROM N.A.
 CC STRAIN=WR-1;
 CC MEDLINE=98105785; PubMed=9445040;
 CC Reddy P.S., Idamakanti N., Zakharchouk A.N., Baxi M.K., Lee J.B.,
 CC Pyne C., Babluk L.A., Tikoo S.K.;
 CC "Nucleotide sequence, genome organization, and transcription map of
 CC bovine adenovirus type 3.";
 CC J. Virol. 72:1394-1402(1998).
 CC [4]
 CC FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
 CC BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC
 CC EMBL: D16839; BAA04115.1;
 CC EMBL: AF030154; BAA09736.1;
 CC InterPro: IPR000931;
 CC InterPro: IPR000939;

```

RL EMBL: AE000291: AAC75061.1: ALT_INIT.
CC -1- FUNCTION: CONTROLS COLONY FORM VARIATION AND AUTOAGGREGATION. MAY
CC FUNCTION: AS AN ADHESIN.
CC -1- SUBUNIT: HETERODIMER OF THE PERIPHERAL MEMBRANE PROTEIN (ALPHA
CC CHAIN) ANCHORED TO THE INTEGRAL OUTER MEMBRANE PROTEIN (BETA
CC CHAIN).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: TO ADHESIN AID-1 AND TO BORDETELLA PERTACTIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000291: AAC75061.1: ALT_INIT.
CC EMBL: D90838: BAA15825.1: ALT_INIT.
CC EMBL: D90839: BAA15832.1: ALT_INIT.
CC EMBL: U24429: AAB47869.1: -.
CC Ecocore: B012686: f1u.
CC Outer membrane: Signal.
CC SIGNAL 1 52
CC FT CHAIN 53 551
CC FT CHAIN 552 1039
CC FT VARIANT 2 2
CC FT VARIANT 41 42
CC FT VARIANT 46 46
CC FT VARIANT 157 157
CC FT VARIANT 188 188
CC FT VARIANT 303 305
CC FT VARIANT 320 320
CC FT VARIANT 372 372
CC FT VARIANT 433 433
CC FT VARIANT 497 497
CC FT VARIANT 585 585
CC FT VARIANT 709 709
CC FT VARIANT 721 721
CC FT VARIANT 751 753
CC FT VARIANT 803 803
CC FT VARIANT 815 815
CC FT VARIANT 824 824
CC FT VARIANT 829 835
CC FT VARIANT 845 847
CC FT VARIANT 855 855
CC FT VARIANT 888 888
CC FT VARIANT 1025 1025
CC FT VARIANT 61 63
CC FT CONFLICT 61 63
CC FT SEQUENCE 1039 AA: 106841 MW: 5170647C08DEBEO CRC64:

```

Query Match 2.9%; Score 147; DB 1: Length 1039;
 Best Local Similarity 20.6%; Pred. No. 0.16;
 Matches 200; Conservative 126; Mismatches 306; Indels 340; Gaps 53;

```

QY 146 AVVKIREGCA--IH-----AQNLYNNHNDVYGMKNFSYRGALISF--ANTFV----- 192
DB 146 AVNTTLNGGGMHMEGAIAGTIND-----KGMQVVKPPTVATDVNTVNGAEGGPD 197
QY 193 SENOSCFLEFMDNICIQTNTAGKGG-----AIYAGTSNFSNNCLF----- 234
DB 198 AEMGDTCGFPRGDAVR-TINKNGRQIVRAEGTANTTVVAGGDDQYHGAALDTTLNGGY 256
QY 235 -FINNACGAGAIIFS---PICSLTGNRGNIVFYNNCFKNVETASSEAS---DGAIR 285
DB 257 QYHNGGTASTADTVVNSGQIVKNGVAGNT-VNCKGRGLVDAGGTATVTLKGGALV 315
QY 286 VTRRLDVTG-NRGRIF-----SDNITKNYGAL-----YAPVTLVDNPTFYINN--- 331
DB 316 TSTAATVGTINRLGAFVEGKADNVYLENGGRDLVLTGHTATNTATRVDDGDTLDRNGGT 375

```

```

QY 332 ---IANNKGAIYIDGTSNSKISADRHAIENENITVNTVANGSTANPERRNAITVA 388
DB 376 ATTSMNGCVLLAD--SGAAVSG-----TRSDGKAFSICGGOADA----- 414
QY 389 SSSGELLGAGSSQNLIFDPIEVSNAGVSFKNKEADDTGSVFSGATVNSADPHORNL 448
DB 415 ---LMEKGSSTL-----NAGDTA-----TDTVNGGLTTAKO- 445
QY 449 QTRPAPLTLSNGFLCIEDHAQLTVAFRQT-----GVSLEGNCALVSCYKNAG 499
DB 446 -GLTACTTLNNG--AIIILSKTYANDTLTREDALLGGSLT-GNSV---EKSGG 498
QY 500 N-SASNASITLKHIGLSS-----ILKSGAETPLW-----VEPTNSNN 539
DB 499 TLTVSNTTLTKRAVNLNECTLLNDSTVTTDVIAGRTALKLTGSLVNGAIDPT---N 554
QY 540 YRADTATFSLD--VKLSLIDY---GNSPYEST-----DLTHALS 576
DB 555 VTLASGATWNIIDNATVQSVVDLSHAGIHTSTRGTKEVPATLKVNKNGONGTISLR 614
QY 577 SOPMLISSEASDNLRSDDMDFSG--LNVPHYG--WGLMTWMAKTQDPPASSATIT 631
DB 615 VRPDMAONNA--DRLVTDGRATGKTLNLVYVANGSASGLATSG--KGIOVVAINGATTE 671
QY 632 DPO-----KANRPHRTLLTWLPAGVYVSP--KHSPLIANTL-----WGNMLATE 676
DB 672 EGAFVQGNRLQAGAFVYSILNRDSESVYRSENAVRAEVLVYASMLTQAMDYDRIVAGSR 731
QY 677 SLK-----NSAEILPSHPFWGITGCGLCMMVYQDPRENNPGRFMRSSGVSAGAIQAQ 729
DB 732 SHDTGVNGENNSVRLS-----IÖGGHLG-----HDNNGGLARGATPSSG----- 771
QY 730 THFSLKFSQTYTKLERYAKNNVSSKNYSKOGEMLSÖEGFLLTKLVGLYSGDHNC 789
DB 772 -----SYGFVRL-----EGLMRTVEAGMSVY--AGVYGAAGHS-- 803
QY 790 HFYTGENTLSQGTFRSQTMGAVFFDLPMKPGSTHILTLAPLGLIYSSHETEVG 849
DB 804 SYVDKDDDSRACTVDDA-----GCLGGLYMLNVH----- 833
QY 850 AYPRSFSTKPTPLINLVPIG---YKGFMTNATORQAM-----TVELAYO 891
DB 834 -----TSSGLMADIVAGGTRHSMKASSDNDNFRARGWGUGLSLETGLPFSTIDNMLE 886
QY 892 PVLVROEPGIAT-----QLASKCIWPGSGSPSSRRHAM--S 925
DB 887 POLQITWGLSLDDGKDNAGYVFGHSGAHOVYRAGTRLSHNDMTGBGT--SSRAPLRDS 945
QY 926 YKISQQTPLSW 937
DB 946 AKHSVSELPVNW 957

```

RESULT 21
 SLAP-CAUCR
 ID SLAP-CAUCR STANDARD: PRT: 1025 AA.
 AC P35828: 046015.
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE S-LAYER PROTEIN (PARACRYSTALLINE SURFACE LAYER PROTEIN).
 GN R5AA.
 OS Caulobacter crescentus.
 OC Bacteria: Proteobacteria: alpha subdivision: Caulobacter group:
 OC Caulobacter.
 OC NCBI_TaxID=76:
 ON (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE=93007489; PubMed=1393820;
 RA Gilchrist A., Fisher J.A., Smit J.K.;
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
 crescentus paracrystalline surface layer protein.";

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:16:45 ; Search time 71.24 seconds
(without alignments)
1587.670 Million cell updates/sec

Title: US-09-677-752-2
Perfect score: 5086
Sequence: 1 MKKAFEEFLIGNSLGARLRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

SPTREMBL-15:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organella:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_unclassified:*

13: sp_vertebrate:*

14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3648.5	71.7	976	2	09PL47
2	1120.5	22.0	938	2	092883
3	1060.5	20.9	952	2	09K2A5
4	1052	20.7	934	2	092882
5	1046	20.6	934	2	09JSE7
6	951.5	18.7	946	2	09RB60
7	948.5	18.6	946	2	092880
8	793	15.6	1025	2	09PL46
9	655.5	12.9	1013	2	084879
10	645	12.7	936	2	092888
11	644	12.7	936	2	09JSA2
12	636	12.5	926	2	09JSA2
13	620.5	12.2	987	2	P71135
14	620	12.2	930	2	09RL45
15	617	12.1	930	2	092393
16	597	11.7	928	2	086164
17	597	11.7	949	2	09K299
18	595	11.7	1016	2	084880
19	587	11.5	947	2	09JSE2

20	586.5	11.5	914	2	086163	086163	chlamydia p
21	586.5	11.5	928	2	09RB65	09RB65	chlamydia p
22	586	11.5	947	2	092813	092813	chlamydia p
23	585.5	11.5	928	2	092398	092398	chlamydia p
24	583	11.5	978	2	092895	092895	chlamydia p
25	583	11.5	978	2	09RB63	09RB63	chlamydia p
26	575	11.3	973	2	092896	092896	chlamydia p
27	575	11.3	995	2	09K2A1	09K2A1	chlamydia p
28	570.5	11.2	922	2	092965	092965	chlamydia p
29	569.5	11.2	922	2	09K1Y9	09K1Y9	chlamydia p
30	568.5	11.2	922	2	0924H9	0924H9	chlamydia p
31	567	11.1	1276	2	09JRM2	09JRM2	chlamydia p
32	564	11.1	983	2	09PL44	09PL44	chlamydia p
33	562	11.0	1407	2	092899	092899	chlamydia p
34	546.5	10.7	986	2	09PUJ3	09PUJ3	chlamydia p
35	544.5	10.7	847	2	P71132	P71132	chlamydia p
36	531	10.4	839	2	P77792	P77792	chlamydia p
37	530.5	10.4	772	2	09RB71	09RB71	chlamydia p
38	530.5	10.4	846	2	P71133	P71133	chlamydia p
39	525.5	10.3	841	2	0923A1	0923A1	chlamydia p
40	519.5	10.2	975	2	084417	084417	chlamydia p
41	517	10.2	1723	2	092812	092812	chlamydia p
42	517	10.2	1723	2	09RB59	09RB59	chlamydia p
43	517	10.2	1732	2	09K2C1	09K2C1	chlamydia p
44	512	10.1	392	2	092881	092881	chlamydia p
45	474.5	9.3	1609	2	09RB58	09RB58	chlamydia p

ALIGNMENTS

RESULT 1

AC 09PL47 PRELIMINARY: PRT: 976 AA.

ID 09PL47:

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.

GN TC0261.

OS Chlamydia muridarum.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=83560;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MOPN / NIGG;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., White O., Hickey E.K., Peterson J., Uterback T., Berry K., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M., "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.";

RT Nucleic Acids Res. 28:1397-1406(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M., Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

RL EMBL: AE002293; AAF39130.1;

DR TIGR: TC0261;

SO SEQUENCE 976 AA; 105964 MW; DBEA56158E3CEEA9 CRC64;

Query Match 71.7%; Score 3648.5; DB 2; Length 976;
Best Local Similarity 70.5%; Pred. No. 9.5e-241;
Matches 689; Conservative 121; Mismatches 154; Indels 13; Gaps 3;

OY 1 MKKAFEEFLIGNSLGARLRFVPSRIFLMPNSVDPFTKESLSKISLTJGDTNHTNCTICYN 60

DB 336 DNGDIFNNNN-----TASKHALNPPRYNAIH-STPMNLIQIGARPGYRVLFYDPIEH 386
QY 413 S-NAGVSVPFKNEADOTGVSFSGATVNSADFHQRNLQTKTPAPLTLNSGFLCIEHQAOL 471
DB 387 ELPSSEPLIFNEGTGTGVLSEGEHQNFTDENMFSLRNTSELRGVLAVERGAGI 446
QY 472 TVNRFQOTGGVAVSLGAVLSCYKKNAGN-----SASNASTILKHIGLNLSTLKS 523
DB 447 ACYKFFQGGTLLGGGAVITTT---AGTIPPTSSPTTGVSTTLNHLAIDLPSLSTQ 502
QY 524 AEIPLWVEPTNNNNYATDPAFESLDVKSLLIDDYGNPSPESTDLTHALSSQPMLEI 583
DB 503 AQAPKIMTPTGTSTYEDSNPTITISG-LLRNSNNEPDYDLSLSEKVLVI 561
QY 584 SEASDNLQASDDMFSGLVN-PIHYGOGIWTGMAKTQDEPPASSATTIDPOK---ANRF 639
DB 562 VVVAQKINSQSLDSTLNGEHYGIGIMSTYVET-----TTITNPTSLGANTR 613
QY 640 HRTLLTLWLPAGYVSPKHSRPLIANTLWGNMLATESLKSNAELTPSDHPFGITGG-- 697
DB 614 HKLLVANNMSPFLGTRPERRGEFTNALMQSATLALAGLSLSS-----WDEKGA 665
QY 698 ---GLGMYVODPRENHGFHNRSSGYSAGMIA--GQTHTESLKSQTYTKLNEYAKN 751
DB 666 ASLOGIGLLVHOKDKNGFGRSHMTGYSATTEATSSQSPNFSLGFAPQFSAKKEHESON 725
QY 752 NWSKRY---SCOGEMFLSLQSGFLTLKVLG-YSYGDHNCHEFTYUGENLTSOGTFRQ 807
DB 726 STSHNHFFSGMCIENTLF--KEWIRLS--VSLAYMFTSEHTHYOGGLEGNOSGCFHNH 781
QY 808 TMGAVFDFLPMKPGSTHILTAFLGALGYSSLHTEVGAYPSPFKTPLINVLVP 867
DB 782 TLGALISCFLPQPHES--LQIYPTTALALRGNLAFQESGDHAEFSLHRLDVSILP 840
QY 868 IGVKGFMAATORPQAMVELAYOPVLVROEPGIATOLLASKGIMFGSSPSSRIHMSYK 927
DB 841 VGRBASMKHNRHRYLWMLVEISRYTLRODPELSKLLISOGTWTQATPYVNNLGIK 900
QY 928 ISOOTPLSMVLHFOYHGFYSSTFCNLTNGEIALRF 965
DB 901 VKNTMQVFPKVTLSLDYSADISSSTLSHYLVNASRMRF 938
RESULT 3
QYK2A5 PRELIMINARY: PRT: 952 AA.
AC QYK2A5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.
GN CP0285.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Uteback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
DR EMBL: AE002190: AAF38142.1: -
DR TIGR: CP0285: -
SQ SEQUENCE 952 AA: 104196 MW: FICB050DA6B74B5A CRC64:

Query Match 20.9%; Score 1060.5; DB 2; Length 952;

Best Local Similarity 30.6%; Pred. No. 4,4e-64;
Matches 303; Conservative 164; Mismatches 447; Indels 77; Gaps 28;

QY 6 FFELIGNSLSGLAREVPSRIFLMPNSVPD-----PTKESLSNK-----ISL 46
DB 8 FLEVLGNFTACMGMTPAYVSLQTDLEKFALEDERFPTSPLLDLSLTGLGFPITTF 67
QY 47 TQDTNULT-NCYLDNLRILALILQKTPNGCAVTTIDYLSFDTQKEGTFPAKMLTPESG 105
DB 68 VGNRHNSQDIYLSNKSIDNILLMTSAGAVSCNPL--LSNVEDHAFESKKNLAIGTG 125
QY 106 GAIGVASPNSPVEIRDTPVIFENNCCRPPTSSNPAANAKIREGAIHAQ--NLYIN 164
DB 126 GATACGG-----ACTTKNKGPLIFESNRGLN-----NASIGETRGALACNGDFTIS 174
QY 165 HNHVVGFKNKFSYVRGAISTANFEVSENOFCLEMDNICIOTNTAGKGAIVAGTSN 224
DB 175 QNCGTFEYFNNSVNMNGALSTNGHCRIOENRAPLLEFN-----NTAPSGGALRENT 228
QY 225 SPFSNNCDLFTINNACCAGATFSP-L-CSLTGKRGIVYNNRCFRNVTTASSESDGA 283
DB 229 TISDNTRPYIFKNNCGNNGAIDQTSVAIKNNSGVIFENNTALSG--SINGNSGGA 286
QY 284 IKVTTFLDVTGNGRIFSDNITKNYGAIYAPVYTLVNGPFYFINTANNKGAIYID 343
DB 287 I-YTNLSTIDNDNGTILFNNNNCIRDGALCTOFLTKNSHYFTNN--QGMNGALMLL 344
QY 344 GTSNKSISADRHAIIFENIVNTNANGTSTANPERRNATIVASSGELLGAGSSON 403
DB 345 QDSTCLLFAEQGIAFQNNNEVLTFTG-----RNALIHCPNS--NLQGANAKVT 393
QY 404 LIFPDTEVSNAGVS--VSRNKEADOTGSYFSGATVNSADFHQRNLQTKTPAPLTLNSGF 462
DB 394 TAEFDPIEHQPTTNPLINPNANMOGITLFSAYIPRESDENNEISSKSTSELRLN 453
QY 463 ICIEDHQAOLTVNFEOTGGVSLGAVLSCYKKNAGNSAS--NASITLHIGLNLSTLKS 521
DB 454 LSTEDRAGNOFKFTQKGLIKGHAASIAITTANSETPTSTVSGOYIINNLIINPLSLA 513
QY 522 SGAELPLWVEPTNNNNYATDPAFESLDVKSLLIDDYGNPSPESTDLTHALSSQPM 581
DB 514 KG-KAPTLMIIRPLQSSAPTEEDNNPTITLSG--VLTJLNEENDPDYDLSLSEPLQNHLL 571
QY 582 SISEASDNLQASDDMFSGLVN-VPHYGOGIWTGMAKTQDEPPASSATTIDPOKANRF 640
DB 572 SLSDVTARHINTDNFHPESLNATEHYGIGIMSPYVET--ITTNNASI---ETANTLY 626
QY 641 RTLLTLWLPAGYVSPKHSRPLIANTLWGNMLATESLKSNAELTPSD--HPFGITGGG 698
DB 627 RALYANMTPLGKRVNFEYQGDALATPLMQSFHTMSLSLNSYNTGSDIEREPLFIOGIA 686
QY 699 LGMVYVODPRENHGFHNRSSGYS--AGMIAGQTHTESLKSQTYTKLNEYAKNNVSSK 756
DB 687 DGIHFVQNSIPGAPGRIGTQSTGYSLQASSETSLQKISLGFAPQFRTMEICSSNNVSAH 746
QY 757 N--YSCOGEMFLSLQSGFLTLKVLGYSYGDHNCHEFTYUGENLTSOGTFRSGTGMGAVF 814
DB 747 NTVSSLYVELPW-FQDAF-ATSTVLAYGQDHLHLSLHPSHOE-OAEGTCYSTHLLAAIG 803
QY 815 FDLPMKPFESTHILTAFLGALGYSSLHTEVGAYPSPFKTPLINVLWLPVIGKGSF 874
DB 804 CSFPMQOKSYLHL--SPYQALAIRSHQTAFAEIGNPKPKFYQAKFFYMLTLPLIGIGW 861
QY 875 MNATORPQAMVELAYOPVLVROEPGIATOLLASKGIMFGSSPSSRIHMSYKISOOTOP 934
DB 862 QSKFHVPTFEMTELSYOPVLVQOONPOIGVTLASGSMIDLGNVRYNALGKVVHNOTAL 921
QY 935 LSWLTLHFOYHGFYSSTFCNLTNGEIALRF 965
DB 922 FRSLDLFLDYOGVSSTSTHHLQAGSTLKF 952

RESULT 4

Db 359 HCTGNS-NLGGANKGYTATFAFDPEHGHPTTNPILFPMNHHOITLFFSSAYIPEADY 417
OY 445 QRIQOTKTPAPLTLNSGLCTIEDHAQTLVNRFTGTGVYSLGNGAVLSQYKNGACNSAS- 503
Db 418 ENNFSSSKNTSELNGSLIEDRAGMOFYKFTOKGILKLGHAASIAITANSETSTSV 477
OY 504 NASITLKIGLNLSSILSGAEIPLWEPPTNSNNYADTAATFASLSLDVKSLTDGCGN 563
Db 478 GSOYIINNLAINLPSILKKG-KAPLMIKPLLOSSAPFTEDNNPTITLSC-PLTLUNEENR 535
OY 564 SPYESTDLTHALSSOPMLISEASDNOLSRDMDPSGLN-VPHYGMOLMTGMNAKTODP 622
Db 536 DPYSDIDSEPLQNHLLSLSDVTAHINTDHFHESLNATEHYGYGIMSYWET--I 593
OY 623 EPASATITDPOKANRFRHTLLTWLPAGYVPSPKHRSPLIANTLGMNMLATESLKMSA 682
Db 594 TTNNAASI--ETANTLRYALYAMNTPLGYKYNPEYOGDLATTPLOSFHTFSLRSYN 650
OY 683 ELTPSD--HFMGITGGGLGMVYODPRENHGFMHRSRGYS--AGMIAGOTHTFSLKFS 738
Db 651 RTGSDIERPPLLEIGADGLFVHONSTIPGAFPRITQSTGYSLOASSETSLHOKISLQPA 710
OY 739 QTYTKLNERAKNNYSKKN--YSCGEMLFSLQEGFLTKLGLYSYGDHCHIFETGGE 796
Db 711 QFFTRKEIGSSNNYSANNTVSSLVELPW-FQEAFF-ATSTYLAAGYGDHHLHSLHPSHQ 768
OY 797 NLTSQGTFRSOTMGAVFDDLPMPKPFSGTHILTAFLGALGYSSLSHTEYGAVPKFS 856
Db 769 E-QAEGTCYSHTLAAICSPFMQOKSYLHL--SPFVOAIAIRSHQTAFFEEIGDNRKFFV 825
OY 857 TKPLINLVLPYGVSGSMNATQRPQATVELAYQPVLYXROPGLATQLLASKGIMFSG 916
Db 826 SOKRPYNLPLPGIGQKMSKFRVPTMTLSTVQPVLYQONPQICVTLASGSGWDLIG 885
OY 917 SPSSHHASYSKISQOTOPLSWLTLFHFYHGFYSSSTFCNYLGEIALRF 965
Db 886 HNYVNALGYKVHNOTALFRSLDLFLDYOGSSVSTSTHHLGASSTLKF 934
RESULT 6
OYRB60
ID OYRB60 PRELIMINARY: PRT: 946 AA.
AC OYRB60:
DT 01-MAY-2000 (TEMBLrel. 13. Created)
DT 01-OCT-2000 (TEMBLrel. 15. Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15. Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN E/F FAMILY (POLYMORPHIC MEMBRANE
DE PROTEIN E/F FAMILY).
GN PMP_18 OR CP0283.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hatori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RA MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Quinn M., Nelson W., Deboy R., Kolonay J.,
RA McLarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).

DR EMBL: AP002546; BAA98677.1; -
DR EMBL: AE002189; AAF38141.1; -
DR TIGR: CP0283; -
SO SEQUENCE 946 AA; 103611 MW; DCBF6B3D04AC452C CRC64;
Query Match 18.7%; Score 951.5; DB 2; Length 946;
Best Local Similarity 29.1%; Pred. No. 1.2e-56;
Matches 281; Conservative 168; Mismatches 416; Indels 101; Gaps 34;
OY 35 PTKESLSKISLTGDTNLT-----NCYLDNLRYLALITOKTPNEGAAYTIDYL 84
Db 47 PLIDITLN---MTPYSHRATLFGVRDITNODIVLDHQSIEMFENFSGDGLSCKS-L 102
OY 85 SFEDTQKEGITYFAKMLTESGAIGYASPNSTYVIRITIGPVIFENNCCRPFTSSAPN 144
Db 103 ALTNT-KNQLIFLNSFAIKRAGAM-YVNGN--FDLSNHOSIIFSGNL-----SFPN 150
OY 145 AA--VNKIREGGAHMOULYINHNHDVYGFKKFYSYVGAISTANTFVSENOGCFEM 202
Db 151 ASNFADTCTGCAVLCSSKVTTSKNOGTAFVINKAKSSGAIQAAIINIKONTGCLFF- 209
OY 203 DNICIQTNAG--KGAIVAGTSNSFESNCDLFFINNACCGAGAI-FSPICSLTGNRGN 259
Db 210 -----NNAGGTAGALFANACR-IENNSQPIYFLNNGSLGGAIRVHOCILTKNKS 262
OY 260 IVFYNNRCFKNVETASSASDGAIKVTRLDVTGNRGRIFPSDNTKNYGAIYAPVYT 319
Db 263 VIFNNN--FAMEADISAHNSGCAIYCS-CSIKDNPGIAFDNNTARDCALCTOSLT 319
OY 320 LYDNGPTFEINNIANNKGAIFYIDGTSNKSISADHAIIFENNTANTNTANGSTSANP 379
Db 320 IODSGPVYFTNN-OCTMGCAIWLRODGACTLPADGCGIIFNNRHKOTFSN--HVSVC 376
OY 380 PRNATVYASSGSEILLGAGSSQNLIFYDPI--EVSNAQSVSFKNEADOTGGSVSG 435
Db 377 TRNVSLTYGASGQ-----HSATFYDPILORTIONS--IQFNNPHELGLIFLSS 425
OY 436 A-----TNSADF--HQRNLOTKTPAPLTLNSGLCTIEDHAQTLVNRFTGTGVYSLGNG 488
Db 426 AYIPDTSTRDGFISHFRN-----HIGLYNGLTALDEKAEWKYKFDQCGTIRLCSR 478
OY 489 AVLSCYKNGAGSASNASITLKLGLNLSSILKSGAEIPLWEPPTNSNNYADTAATF 548
Db 479 AVFTTTEDEQSSSVGSVINNNLAINLPSIL-GRRVAPKAIPTGSGAPYSEDNPII 537
OY 549 SLSDVKSLIIDYGNPSPESTDLTHALSSOPMLISEASDNOLSRDMDPSGLN-VPHYG 607
Db 538 NLSG-PLSLDDENLDIPYDADLAPPIAEVPLLYLDVTAHINTDNYFPEGINTQHYG 596
OY 608 WGLMTWGMAKTODPEPASSATITDPOKANRFRHTLLTWLPAGYVPSPKHRSPLIANTL 667
Db 597 YGVWSPYWIETITTSSTSS-----EDTVNLTLRHOLYGDMPPTYGVNPEKNGDIALSAF 651
OY 668 W----GNMLLATESLKNSAEITPSDHPFGITGGGLGMVYODPRENHGFMHRSRGYSAG 724
Db 652 WOSFHNLFATLRYOTQOQIAPT-----ASGEATRLVHONSNDAAGFHEATGYSLG 705
OY 725 MA--GQHTFSLKFSQTYTKLNERAKNNYSKYSOCGEMLFSLQEGFLTLTLYLV 781
Db 706 TTSNTASNSHSEGVNFSOLFNSLYESHSDNSVASHTTYALQIINNWLDERSTASLSA-Y 764
OY 782 SYGDHCHHFFYTQGEN--LTSQGTFRSOTMGAVFDDLPMPKPFSGTHILTAFLGALGY 839
Db 765 SVSN--HHIKASGYSGKIOTEGCKYSTLLGALSLSLQ-WRSRPLHFTPEIDAIYR 820
OY 840 SLSLHTEYGAVPKFSRSTKPLINLVLPYGVSGSMNATQRPQATVELAYQPVLYXREP 899
Db 821 NQOTAFQESGDARFVHKPLYNLTPLGLOSARESKFRPLTYVNIETLAPVLYQONP 880
OY 900 GIATOLLASKGIMFSGSPSSHHASYSKISQOTOPLSWLTLFHFYHGFYSSSTFCNYLNG 959
Db 881 EVNVSLESSESSGMLSLSTLARNALIAFGKRNQIIFPRLSVFLDYOGSSVSTSTTHHYHA 940

```
Db 80 LSTNSSGAFV--ESMTSETAIDNADLFECCNNYCTHOG--GGAINATGLISFKNQNIL 135
QY 128 IFENNTCCRPFTSSNPMAVNIREGGAIHAQNLVYNNHDDVVGPKNRSYRGAGISIA 187
Db 136 FNNNTTIGTFGVALTRERNR--GGALYSSIEILNHS--LNFNNTSGMGAGVSTI 191
QY 188 NTFVSENOCEFLFMDNICIQ--TNT---AGKGAIVAGTSNSESFNCDLFEFFINACC 241
Db 192 QNLVYIKNTSGIYAFENNHHTDHPNFATILARGAGVGGQAGCEISHNTGPVFNNTYCG 251
QY 242 AGCAIIFS-PICSLTGNRNIYVYNNCFKNVETASSEASDGAIKVYTRLDVYGNCRIF 300
Db 252 YGCAISTGQCIJFRDKDKLFIINNALSQMHNT--SAQNGAVISAGEFGLLNKGPYI 309
QY 301 FEDNITKNGCAIYAVVTVLVNDGPTTFINNANKGAIYIDGT-SNKKISADRAIIF 359
Db 310 FENNMAVYIAGAISSCNMLFQENGPYFLNNSA-LYGGAFHLFASPAANYIHTGSDIIF 368
QY 360 NENITVNTNAN-----GTSTANPPRRNATTVASSGCEILLGSSQNLIEYDP 409
Db 369 NNNTLSTTGMSAGLRKLFYIPGTINN-----NPITLS-----LGAKKOTRIYFDL 415
QY 410 IE-----VSNAVSVSEFKREADQTSVFGATVNSADFHORNQOTKT-----PA 454
Db 416 FQMGGLKANTRPPENSPHTVTINPDEFGAVVFSYKNI-SSDLQAHIMIASKTHNQIKDS 474
QY 455 PTLISNGFLCTIDHAQ-----IYVNRFTQTCGVVSLCNCGLAVSCYKKGAGNSASNAI 507
Db 475 PPTLFGTMSIENGAEFEFFNGPLT---QESTSLALODSDILTVGK-----DASL 522
QY 508 TLKHGLNLSLTK-SGAEIPLWVEP---TNNSNYTDATATSLSDVKL-----SL 557
Db 523 TITHGLIILPGLINOOGTAPRIARVNPDMOTONTNSNOA-PVSTENAVQOKLFFSGSL 581
QY 558 IDDYGNSPYESTDLTHALSSQPMLSISEASDQNRSD--DMDFGLNVPYHGMOLMTW 614
Db 582 VDENEYEVSDCLSRKANOPILHETINDAQLSDMKMTINTLSLPHYGYGLWTS 641
QY 615 GWAKTQDPPASATITDPOKANR-----638
Db 642 NMMTT--TRTVSLJNSTEQTANNNSIQEOKNTSETFDSNSTTAKLPSIRASTGCTPLA 699
QY 639 -----FHRTLLTWLPAGVYSPKHSPLIANTLWGNMLATESKNSAEITPSDHPW 692
Db 700 TTDVTVTRISLAVSWPTIGYIADPARCGLIANNIYSSCRNTTLVLR--LLPDD-SWF 755
QY 693 GITGGGLGMVYQDPRNHPGFHMSSGY--SAGMIAGOTHTFSLFSQTYTKLNERVAK 750
Db 756 ALQGSATLFTKQKRLDYHGYSASKGYAISQASGAIGHKFLFSQSDTMMEKRTN 815
QY 751 NNVSCKNYSCQCEMLFSIQEGFLTKLVG--LYSGDNHCHHFTYOGEMLTQCGTFRSOT 808
Db 816 NMISSRYYL---SALCFEPMEDRIALLGAAYVNGTHTYNY--GTYKFKFGNHSTT 870
QY 809 MCGAIFFDL-PKMPGSGTILTAPELGAIGYSSLSHFTFEVAVPRPSFTKPLIIVLV 867
Db 871 LGGSALRCELRDMPFOS--IMLTPICALISRTPEASIOEGDLARIFELKOPHAYVSP 928
QY 866 IGVKGSFNATQRPQAWVELAYQVLYRQEPGIATOLLASKGIFWGSQSPSSRHAMSYK 927
Db 929 IGIKGYSSNKKPJVSCHEVAVOPTLLYWRPILNTVLKKNKSMETNTPLAKHSF-YG 987
QY 928 ISOQTOPLSMLTHFOYHGFYSSSTFCNVLNCEIALRF 965
Db 988 KGSLSLKFSYLLKLPANYQAVATSTVSHYNAGALVF 1025
```

```
RESULT 9
ID 084879 PRELIMINARY: PRT: 1013 AA.
AC 084879:
DT 01-NOV-1998 (TREMblrel. 08, Created)
01-NOV-1998 (TREMblrel. 08, Last sequence update)
```

```
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN G.
GN PMPC.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RX MEDLINE=99000809; Pubmed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: AB001360; AAC68469.1;
SQ SEQUENCE 1013 AA; 107366 MW; F0927743C0A651DD CRC64;

Query Match 12.9%; Score 655.5; DB 2; Length 1013;
Best Local Similarity 25.7%; Pred. No. 2.1e-36;
Matches 261; Conservative 152; Mismatches 388; Indels 213; Gaps 43;

QY 39 SLNKRISLTGDTNHLNCLYDLNLRILALLOKTPNDEGAVITTDYLSFEDTQKEGIFAK 98
Db 84 NIIAGSTVYIGRCHSLT--FENIR-----TSTNGAALS-----DSANGLF--- 121
QY 99 NITPESGCAIGYASNPSPVEIRDTIGPVIFENNCCRPFTSSNP-NAV----- 147
Db 122 --TIEFKLSTSNQNSLAVL-----PATTNNSQPTTSTSTNSGITYKTDLLLLL 174
QY 148 -----NKIR-FEGGAIHAQNLVYNNHDDVVGPKNRSYRGAGIASTANTFVSENO 198
Db 175 NEKFSFYSNLVSGDGAIDAKSLTVQGISKLQVFOENTAOADGACQVVTSPSAMANAP 234
QY 199 FLEMONICQIQTNTAG-KGGAIIYA-----GTSNSE-----SNCDLFEINACGAG 243
Db 235 IAVTANV-----AGVRGGGIAVODGCGGVSSSTEDPVVFSFRNIAVERDGVNARVG 288
QY 244 GAIFSPICSLTGNKGNIVFYNN--RCFKN-----VETASSEASDGAII 284
Db 289 GGIYS-----YGVAVFLNNGKTLFLNNVASPYIAEOPTNGQASVTSQNYDDGAI 340
QY 285 KYTRLDVTGN-----RCRIEFSQNTIKKNGAIYAVPVTLVNDGPTTFINNIA 334
Db 341 PCKNGAAGSNNSGSVSFDEGVVFFSSVAAGKAIYAKKLSVANCQPOVQGLNIA 400
QY 335 NKGAIIYIDGTSNKSISADRHAIITENET---VNVYTNANTSTISANPPRRNATV 390
Db 401 D-GGAIYVIGSESLASAVYGDIFDGNLKRTRAKENAAADVNAVTVSS---QALSKSG 454
QY 391 SGEILLGAGSSONLIFYDPIEYVN-----AGVSYFKEKEDQTSVYFGATVNSADF 443
Db 455 GKITTLRAAGHQIILFNDPIEMANGNQPQSSSEPLKINDGCGYGDIVF--ANQNSTLY 512
QY 444 HORNILOTKTPAPLITISNGFLCTIEDHAQLTVNRFOTQGVVSLCNCGLAVSCYKNGAGNS- 501
Db 513 QN-----VTEOGRIVYLRERAKLSVNSLSOTGGGLYMEASSTIDFVPPQPOPP 562
QY 502 ASNASITLKHIGLNLSTLSKSGAEIPLWVEPTN--NSNNYVADIAAFPSLDVLS- 556
Db 563 AANOLITLSNLHLSLSLANNNA-----VNPPTNPPAODSHPAIIIGST--TAGSVTISGP 617
QY 557 LIDDYGNSPYESTD-----LTHALSSQPMLSISEASDQNRSDMDPFGSLNVPHY 606
Db 618 FREDLDDTAYDRYDMLGNSQKIDVLKLDGTOP--SANAPSDLT-----GNEMKY 667
QY 607 GWQGLTWGMMAKTQDPEPASATITDPOKANRPHRTLLTWLPAGYVSPKHSPLIANT 666
Db 668 GYGGSNKLAW-----DPTANNNGPYTLKATWTKTGYNCPERVAVSLVPS 712
QY 667 LMGNNMLATESLKNSEAEITPSDHPW-GITGGGLGMVYQDPRNHPGFHMSSQSYSGM 725
```

RA Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
RA McClarty G., Salberg S.L., Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546: BAA98653.1;
DR EMBL: AE002193: AAF38165.1;
DR TIGR: CP0308;
SQ SEQUENCE 936 AA: 100105 MW: 3981DB3C950AF95A CRC64;

Query Match 12.7%; Score 644; DB 2; Length 936;
Best Local Similarity 27.0%; Pred. No. 12e-35;
Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

OY 75 GAAVTTTDLSEFDKQEGIVFAKNTLPESSGAIGY-----ASPSPTVEIR 121
DB 54 GTTYSLSLDVDFQNGALGIPLAGSGFLBAGDLPFGQNHAKLFAFINAGSAGTAVST 113
OY 122 DTICPVIFENN-----TCCRPFSSNMAAVNKIREGALHAOMLYINHHHDVYGFMK 174
DB 114 SAADKNLLFNDPSRLSIISCPILLSPTGQCALKSV-----GSLSTGNSQII-FTQ 164
OY 175 NFSYVGAISTANFVSENOGSCFLMDNICIOTNTACKGAIYA-GTNSPESNCDL 233
DB 165 NFSSDNGVYINRKN-FLISGTSQFASFNRN---QATGKGGGVYATGTTTENSFGIYS 220
OY 234 FEINNAACGAGAFS-PICSLTGNRCNIVFYNNRCFKNVETASSSEASDGAIKVTT--R 289
DB 221 FSONLAKGSGALYSTDNCSITDN-FQVIFDGSAMEAQAQ--ACGALCCTTDTKT 273
OY 290 LDVTGRCGIFESDNTKRYGAIYAPVYTLVDCNGPTFYFINNANK-----GCAITYDG 344
DB 274 YLTGTGRK-NLSTNTNLTALYGAISGLKYSISAGGPTLFQSNISGSSAGGCGALINIAS 332
OY 345 TNSKRSADRAHAIIFENIVTNTNANGTSTSNAPPRRAITVASSSGEILGAGSSONL 404
DB 333 ADELALSATSGDITFNNQVTN---GSTST-----RAHINIDTAKTTSIRATGOSI 382
OY 405 IFYDPIEVSNAGVS-----FNKEDQTSVSESGATVNSAD-FHORNLOTKTP 453
DB 383 YEYDPI-TNPGTAATDITLNLADANSEIEYGAIVSEKSLPTKAIANVTSTIR 440
OY 454 ABLTSLNGLCIEDHQLVNRFTOTGCVVSLGNCALVISCYKNGAGNSASITLKHIG 513
DB 441 QPAVLARGLVLRDGVTFEFDLTOSPGSRILMDG-----GTLTSAKEANLSINGLA 492
OY 514 LNLSSILKSGAEIPLVMEPTNNSNNTVADTATFSLSDVKLSLIDYGNPSPESTDLTH 573
DB 493 VLSSL-----DGTNKAALTEADKNISLST-TALIDTEG-STENNNILMS 538
OY 574 ALSOPMLTISEASDN-QLRSDMDSEGLNVP--HYGMOGLMTWGNKATODEPASSATI 630
DB 539 A-STYPLELTLTGAAGNLTTLGALSTLTQEPETHGYGQNMQLSMAN-----ATSSKI 591
OY 631 TDPQANRHRKLLTLPLAGYVPSPKHRSPLANTLKGNNMLATESLANSALTELPSSHP 690
DB 592 GS-----INMTRTGYIPSPERSNPLPLNSLMGN-FIDIRISINQLIETKSSGEP 638
OY 691 F-----WGITGGGLGMVYODPRENHGCFHMRSSGYSAGIACOTHTFSLKFSOTYKLE 746
DB 639 FERELM-LSGIANTFYDSMPTRHGFRHISGVALGITATTAPADQLTFA-FCQLFA 693
OY 747 RYAKNNVSSKN-----YSCQGEMLFSLQEGFL-----LTKLVGL----- 780
DB 694 R-DRNHITKNGDRTGAGALYFHHTGELFDI-ANFLMGKATRAPWVLSLSQIILPSFDA 751
OY 781 ---YSGDHNCNHFFYTOGENTLSOGTFRSOTMGAFFDLPKMPSTSHL--TAFPLA 835
DB 752 KRSYLTLDHNMKTYT--DNSIITKGSWRNDAFCADAGSLPF-VISVPYLKEVEFEVAV 808
OY 836 LCIYSSLSHFTVGAVPASFSTKTPLINLVPIGVKGSFNNATQRQAQMTVELAYOPV- 894

DB 809 QYIYAHQODFYERYAEGRAF-N-KSELINVEIPIGVT---FERDSKSEKTYDULITMIL 864
OY 895 -YRQEGITOLLASKGIFGSGSPSSRHAMSKYISQOTPLSLWTLHFQYHGFYSSSTF 953
DB 865 AVRNRKPCOTSLIASDANMAYGTNLAROGFSVRANHFQVNPHEIFGOF-AFEVRSS 923
OY 954 CYNUNGEIALRF 965
DB 924 RNT-NTNLGSKF 934

RESULT 12
P71135
ID P71135 PRELIMINARY: PRT: 926 AA.

AC P71135;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U72499; A818188.1;
SQ SEQUENCE 926 AA: 98439 MW: 3E75E52F594750F CRC64;

Query Match 12.5%; Score 636; DB 2; Length 926;
Best Local Similarity 24.0%; Pred. No. 4e-35;
Matches 241; Conservative 157; Mismatches 365; Indels 240; Gaps 41;

OY 41 SNKISLTGTHNL--TNCYLDNLRITLALQTPNCGA-----AVTIDY--L 84
DB 82 ADNLTRKGNHSLITNA-----NAGAPRAGIANTADKILTLIDFSKL 125
OY 85 SFEDTQKEGIVFAKNTLPESSGAIGYASPSPTVEIRDTIGVFEENNTCCRPFTSSNPN 144
DB 126 SFECRSSLVNIGKG-AMSGALNANNAS-----ILPDQN-----YSKEN-- 166
OY 145 AAVNKIREGALHAOMLYINHHHDVYGFKNFSYRGCALSTANTFVSENOGSCFLPMDN 204
DB 167 -----GGAISCKAFSLTSGSKSEISFTTNSAKKGAIATGIAHLSDNOGITRFSGN 218
OY 205 ICITOTACKGAIYACGTNSPESNCDLFTINMCC-----AGCAIFPISLTGNRCN 259
DB 219 TAVNS-----GCAVVSSEASMTIAGNN-HVAFSNNAVSGSSDCCGAIH---CSKTGSAPT 269
OY 260 IVEYNNRCFKNVETASSSEASDGAIKVTTRLDVTGNRGRIFESDNTKRYGAIYAPVY 319
DB 270 LTIROKKV-----LIEENTSSAKGAIITDKLI 298
OY 320 LVNCGPTFTNN--IANKGCAIYIDGTSNKSISADRAHAIIFENIVTNTNANGTST 376
DB 299 LTSGGPTAFINNKVTHATPKGAIAGIANGECSLTAEHDITFDNNLMATODNAT----- 353
OY 377 ANPRRNATIVASSSGEILLAGSSONLIFYPDIEV-SNAGVSYSFNK-EADOT--GSVY 432
DB 354 ---IKRNAINIEGNCGFVLRASGKTTFSPYDITVEGNAADLTLNKAEGDGTINGRTII 410
OY 433 FSGATVNS-----ADFHORNLOTKTPAPLTLSNGFLCIEDHQLVNRFTOTGCVVSLGN 487
DB 411 FSGEKLTREQAAVAD---NLKTTFTQPTITLAAGELVLSGVEEAKTVVYVAGSLILMD 466
OY 488 GAVLSCTYKNGAGNSASNASITLKHIGLNLST-----LKSQAEIPLLM-----VEP 533
DB 467 A-----GRLSARTEDATLTNLAINPNTLDGKKFAVAVDAVAGKNVTLGSGAIVDP 518
OY 534 TNN--SNNTADTAATFSLSDVKLSLIDYGNPSPESTDL-THALSSQMLTISEASDNO 590

Db 341 -----TTGSTDTPKRNAINIGSNGKFTLBRANKHTIFFEYDPTSEGTSSDVLKINNGSA 395
QY 426 -----DQTSVVFSGATVNSADPH-QRNIOTKTPAPLITLNSGCLIEDHQAOLTVNFTQ-T 479
Db 336 GALNDYQGTILFSGEITLVADELKVDNLSKSTFQVSLSGKLLLOKGVYLTSTSFSDRA 455
QY 480 GGVNSLGNCAVLSCKYKNGSNGSASNAITLKHIGLNLSTLSKGAIEPLLWVEPTNNSNN 539
Db 456 GSLGMDSGTTL-----STAGSITITNLGINVDSL--GLKQPY-----S 493
QY 540 YTAOTAAFTSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASDNLQSDMDPS 599
Db 494 LTAKGASNKVIYSGKLNLDIEGNI-YES---HMFSDQLFSLKITYDADVDTNVDIS 548
QY 600 GL-NYP-----HYGMOGLMTGMAKTODPEPSSATITIDPOKANRFRHTLLTLPLAGY 652
Db 549 SLIPVPAEDPNSEYGFQGMVNW-----TTDTATNTKEAT-----ATWTKGTF 592
QY 653 VPSPKHRSPLIANTLGMNLLATESLKNASAL--TPSDHP--FMGITGGGGMVYODPR 708
Db 593 VPSPERKSAIVCNTLIMG-VFTDIRSLQOLVEIGATGMEHKGQFW---VSSMTNPLHKTGD 648
QY 709 ENHDPGFHMRSSGYSAGMIAGOTHT-----FSLKFSQTYTKLMERY-AKNN----- 752
Db 649 ENRKGFRHTSGY---VIGSAHTPKDLETFPACHLFARBDKCFIAHNSRTYGTLEF 705
QY 753 -----VSSKNYSCOGEMFLSLO--EGF-----LTKLVGLSYGCHNCHHTQCENLTS 800
Db 706 KHSHTLQDPNTLGRARFSESATIEKFPREIPLADVOVSFSHSDNRMETHTYSLE--S 763
QY 801 QGTFRSQTMGAVFEDLPKPPGSTHILTAFLGALG---IYSSLSHTEVAGYRST 857
Db 764 EGSMSNECIAGIGIDLDP-VLSNHPLEKFTIPQMKVEMVYVSONSFESSDORGSFI 822
QY 858 KTPLLNVLVPIGVKVSFNATQPOAWVELA--YQPVLYROEPIATQLASKIWFCS 915
Db 823 GR-LNLSTIPVAK--FVOG-DIGDSTYDLSGFFVSDVYRNPNPOSTATLVMSPSWKIR 878
QY 916 GSPSSRHA 923
Db 879 GGNLSROA 886

RESULT 17
Q9K299
ID Q9K299 PRELIMINARY; PRT: 949 AA.
AC Q9K299
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0302.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. *Chlamydia*.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craen B.,
RA Bowman C., Dodson R., Guim W., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002192; AAF38159.1; -
DR TIGR: CP0302;
SQ SEQUENCE 949 AA: 101357 MW: A00809E16C699BE3 CRC64;

Query Match 11.7%; Score 597; DB 2; Length 949;

Best Local Similarity 26.1%; Pred. No. 1,9e-32;
Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;
QY 36 TRESLSNKLSTLGD-----THNLNLCY---LDNLRYI-----LALQKTPNNGA 76
Db 66 TPRTSATYTLTGDFVEYEPKGTPLSDSCFKQTDNLTLFLNGHSLTFGFIADCTHAGA 125
QY 77 ATTIDYLSFFTOVEGIFAFKNTLPESGAGIYASPNPTYEIDTICPVYFENNCTCR 136
Db 126 AASTT-----ANKNLTFSGFSLTSFSSPSPTV---TTGQ----- 157
QY 137 PTTSSNPNAVKKIREGAIHAONLYINHHNDVCFMKFESVRCGAISTANTFVSENO 196
Db 158 -----GTLSSAGVNLNT---KLVYAG--NFTSADGAIKGA-SFLITGTS 199
QY 197 SCFLFMDNICIOTNPAKGAIVACTSNSFESNCDL--FFINMACAGAIFFPICSLT 254
Db 200 GDALFSNN-----SSSTKGAI-ATTAGARIANNNGYVFLSNIASTSGAIDDEGTSIL 253
QY 255 GNRGNIYFNNRCFKNVERASSEASDGAIKYTTRLD---VTGNRGRFFSDNTTKYNG 310
Db 254 SNKKFLYFEGN-----AAKTGGAICTKASGSPELLISNKKTLIFASNAETSG 303
QY 311 GAIVAPVTVLVNCPITYFI-NNI--ANNKGAIVYDGTNSKISADRAHIIENYIVTV 367
Db 304 GAIIHAKKIALSSGFTTEFLRNVSATPKGAIISIDASELSAETGNITTVRNTLT-- 361
QY 368 TIANGTSTANPPRRNATVASSSGEILLAGSSONLIFYDPI--EVSNAGVSVSFKEA 425
Db 362 -----TTGSTDTPKRNAINIGSNGKFTLBRANKHTIFFEYDPTSEGTSSDVLKINNGSA 416
QY 426 -----DQTSVVFSGATVNSADPH-QRNIOTKTPAPLITLNSGCLIEDHQAOLTVNFTQ-T 479
Db 417 GALNDYQGTILFSGEITLVADELKVDNLSKSTFQVSLSGKLLLOKGVYLTSTSFSDRA 476
QY 480 GGVNSLGNCAVLSCKYKNGSNGSASNAITLKHIGLNLSTLSKGAIEPLLWVEPTNNSNN 539
Db 477 GSLGMDSGTTL-----STAGSITITNLGINVDSL--GLKQPY-----S 514
QY 540 YTAOTAAFTSLSDVKLSLIDYGNSPYESTDLTHALSSQPMLSISEASDNLQSDMDPS 599
Db 515 LTAKGASNKVIYSGKLNLDIEGNI-YES---HMFSDQLFSLKITYDADVDTNVDIS 569
QY 600 GL-NYP-----HYGMOGLMTGMAKTODPEPSSATITIDPOKANRFRHTLLTLPLAGY 652
Db 570 SLIPVPAEDPNSEYGFQGMVNW-----TTDTATNTKEAT-----ATWTKGTF 613
QY 653 VPSPKHRSPLIANTLGMNLLATESLKNASAL--TPSDHP--FMGITGGGGMVYODPR 708
Db 614 VPSPERKSAIVCNTLIMG-VFTDIRSLQOLVEIGATGMEHKGQFW---VSSMTNPLHKTGD 669
QY 709 ENHDPGFHMRSSGYSAGMIAGOTHT-----FSLKFSQTYTKLMERY-AKNN----- 752
Db 670 ENRKGFRHTSGY---VIGSAHTPKDLETFPACHLFARBDKCFIAHNSRTYGTLEF 726
QY 753 -----VSSKNYSCOGEMFLSLO--EGF-----LTKLVGLSYGCHNCHHTQCENLTS 800
Db 727 KHSHTLQDPNTLGRARFSESATIEKFPREIPLADVOVSFSHSDNRMETHTYSLE--S 784
QY 801 QGTFRSQTMGAVFEDLPKPPGSTHILTAFLGALG---IYSSLSHTEVAGYRST 857
Db 765 EGSMSNECIAGIGIDLDP-VLSNHPLEKFTIPQMKVEMVYVSONSFESSDORGSFI 843
QY 858 KTPLLNVLVPIGVKVSFNATQPOAWVELA--YQPVLYROEPIATQLASKIWFCS 915
Db 844 GR-LNLSTIPVAK--FVOG-DIGDSTYDLSGFFVSDVYRNPNPOSTATLVMSPSWKIR 899
QY 916 GSPSSRHA 923
Db 900 GGNLSROA 907

RESULT 18

01-OCT-2000 (Tremblrel. 15, last annotation update)
 DE POLYMORPHIC MEMBRANE PROTEIN A FAMILY.
 GN PMP_19 OR CP0213.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CM1029;
 RX MEDLINE=9920606; PubMed=10192288;
 RA Kallan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RA Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,
 RA Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H.,
 RA Matsushima H., Tanaka C., Furukawa S., Miura K., Nakazawa A.,
 RA Ishii K., Shiba T., Hattori M., Kohara S.;
 RT "Comparison of outer membrane protein genes omp and pmp in the whole
 genome sequences of Chlamydia pneumoniae isolates from Japan and US.";
 RL Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dooson R., Quinn M., Nelson W., Deboy R., Kolonay J.,
 RA McCarty G., Salberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE001638; AAD18679.1;
 DR EMBL: AB033816; BAA85967.1;
 DR EMBL: AE002181; AAF38083.1;
 DR HSSP: Q90121; IKPT.
 DR TIGR: CP0213;
 SQ SEQUENCE 947 AA; 103642 MW; 20CE1DEE1606DFF CRC64;

Query Match 11.5%; Score 586; DB 2; Length 947;
 Best Local Similarity 25.2%; Pred. No. 1, 1e-31;
 Matches 250; Conservative 173; Mismatches 390; Indels 178; Gaps 47;

14 LSGLAKEVPSRFLFMPNSVDPPTKESLSNKT---SLTGDTNLTNCLYMLNRYIIAIILOK 70
 DB 30 LSGHSGEDLEFLTLSSSPRTYTLRLKDFVCFDAGNSHKPGAAFLNLGDLFFINS 89
 QY 71 TPNEGAAVITITDYLSPFDQKEGIYFAKNLTPESGAIYVASPNPTVE-----IRDT 123
 DB 90 TPL--AALTFKNI--HLGARGAGLSESNT-----FKGLSHLVLENMSMGVLTIT 137
 QY 124 IGPVIFENNT--CCRPFTSSNPNAVNKIREGAIHAQN---LYINHHVYGFEMKNF 176
 DB 138 SGDLSEFINNTSYLVCNNMISYGP-----GGALLQGRKSKALFFRDNRGITFLKNK 188
 QY 177 SYVR-----GCAISTATFFVSENGSCFLMDN--ICIQNTNAGCAIYAGT--SNF 226
 DB 189 AVNQEASHGCGAVSSISP-----GSPITFADMOELFQENEGELGAIINDOGAIF 242
 QY 227 ESNNGDLFEINNACGAGIFSPICSLTGNRCNIVFYNNRCFKNVETASSEASDGAIV 286
 DB 243 ENNFQTFSTFSKAKSGAVSRICNLYSOMKDTLF-----TKNAAKVGCAIT- 291
 QY 287 TTRLDVTGNRGIFTSNDITNKYGAIVAVVTLVD--NGPTFYINIA--NKGAIYIDG 344
 DB 292 ADVVHIRDCKGIVFEEN--SATAGIAVNAVCIDINAGPVEFINNSALGLNGAIYMQA 350
 QY 345 T-SNKSISADRAIIFNENIVTNV--TNANGTSTSANPRRAIYVASSSGELLGAGSS 401

DB 351 TGSILRLHANOQDIEFCGNKVRKQFHSNINSTNFTN---NAITIQAPREFSLSANEG 406
 QY 402 QNLIFFDPI-----EVSNAGVSFENKEADQTSVSGATVSADPHONLQTKP---A 454
 DB 407 HRCIEFDPIISATENYNS--LYINHRLLEAGCAVIFSGARLSE--HKENKNKTSIINO 463
 QY 455 PTLTSGFLICIEDHAQLTVNRFOTGGVYSLGNAVLSGCKNAGCAGNSASNAITLKIGL 514
 DB 464 PVLKSGVLSIEGAILAVRSEFYQEGGLALPGSKULT---QGKNSKDKIVITNLGF 519
 QY 515 NLSSILKSG-AEPLILVMEPTNNSNNTYDATTATPSLSYKL-----SLID--DYGNSP 565
 DB 520 NLENLDSDPAET-----RATEKASIEISGVPRVYCHTESFYEYHEHASKP 565
 QY 566 YESTDLTHALSSQPLISEASDNLRS---DDMFSGLVNPHYGMGLMTGMKATDP 622
 DB 566 YTTSTII--LSAKKLVLTARSREKQIQNLIASEYMG-----YGOGSMEFSNBNPTK 617
 QY 623 EPASSATITDPOKANRFRHTLLTLWLPAG--YVSPKRRSPLIANTLGMNL-----LA 674
 DB 618 E-----KRTIIASMTPTGEFSLDPRKRSFIPTTLMSFSGLIANSIV 661
 QY 675 TESIKNSALITSDHPFNGITGGGLGMMYQDPREHHPFHRSSGYSGMIAGOTHTFS 734
 DB 662 NNHYLNNESEVTPLOH--LCVFGPVPYQIMEONPKOSSNNLVOHAGHNVC--ARIPIFSFN 717
 QY 735 LKFSQYTKLNERYAKNNYSKNYSCOGEMLF---SLOGFLTKLVGLYSYGDH---C 788
 DB 718 TILSALITLFFSSSQQVADASHA---QILGYTSLNKSQALSRSSFSTEDSOYWK 774
 QY 789 HHFYTGEMLTSGCTFRSOTMGCAVFFDLPMKPFQSTHILTAFLGALCIYSSL--SHT 846
 DB 775 HVFPYKG---TSRGSWRNMGSGVSMYAV--PKGIRYLKMPFDLO--YTKLVQNPV 828
 QY 847 EVCAYRSTSTKPLINLVPIGVKGSFNNATQRPDAW--TVELATQPVLYROEPIATOL 905
 DB 829 ETGYDPRYFS--SEMTNLSLPIGIALEMRPFGSRSLFLQVSTSYIKDLRRVPOSSASL 887
 QY 906 LASKGIFGSGSPSRHAMS-----YKI 928
 DB 888 VLNHYTWMDIQGVPLCKEALNITLNSTIKYKI 918

RESULT 23
 Q92398
 ID 092398 PRELIMINARY: PRT: 928 AA.
 AC Q92398;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G
 CN OMP10 OR PMP_9 OR CP0306.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CM1029;
 RX MEDLINE=9920606; PubMed=10192288;
 RA Kallan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VR1310;
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christiansen G., Birkelund S.;
 RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily
 of autotransporting pathogenicity factors.";
 RL Submitted (Jan-1999) to the EMBL/Genbank/DBJ databases.

QY 154 G-----AIIAONLXIIHNHDVYGFMKNFSGVYGGAIATANTFVVS--ENOSCFLEMDN 204
 DB 165 ASALOPTDSLTVEN--ISOSIKIFGFLNLF-----GSAISSSTAVAKFINNTATMSFSHN 218
 QY 205 ICIOTNTAGKCAIYAGTSNFSFNCDLFFINNACC-----AGCAIFS 248
 DB 219 F-----TSSGGVYIGGSLFENNNGCIIFTANCVSLSKGVTPSSGTALGSGAICI 273
 QY 249 P--ICSLTGNRCNIYFVNNRCFKVETASSEASDCAIKVTRLDVTGNRCGRIFESDNT 306
 DB 274 PTGTFFELKNNGCKCTFSYN-----GTPNDAGAIYAEI-CNIVNOCALLDSDNTA 322
 QY 307 KNYGAIYAPVYTLVDNCPTEYFINNANKGAIYI-----DGTNSKISADRHAI 358
 DB 323 ARNGAICAKVANIIOGRPIESRRRA-EKGAIFIGSVGDPKAKOTSTLTLLASEGDI 381
 QY 359 FENIIVTNTANGSTSTANPPRRNAITVASSGCEILGAGSSONLIFYDPI-----E 411
 DB 382 FCGNML-----NTKPGIRNAITVEAGGEIVLSAOGGSLRYDPTTHSLPTTS 430
 QY 412 VSNAGSVSFNKEADQTSVVFSGATVNSAD-FHQRNLQTKTPAPLTLNSGFLCIEDHAQ 470
 DB 431 PSNKDITINAN---GASGVVFTSKGLSTELLIPANTTTLLGTGVKIASGELKITDNAY 487
 QY 471 LTVNRF-TOTGGVSLGNGAVLSCYKNGAGNSASNAITLKHIGLNSLSKSGEIFLL 529
 DB 488 VNVLFATOGSGQLTLGSGGTGLATPTGAPAA--VDFITGLADPSPFLKRD----- 539
 QY 530 WVEPTNNSNTYADTAATFSLSDVLSLIDYGNSPYESTDLTHALSSQPMLSISEASDN 589
 DB 540 FVSASVANG--TKNVTLTGAL-----VLDE-----HDVTDLYDNVSLQSPVAIPAIY-- 584
 QY 590 QLRSDMDPSGLN-----PHYGQGLMTGMAK-TDPEPASSATITDPOK 635
 DB 585 -----FKGATVTKTGPDEIATPSHYGCKMSYTWSPHLLIPADGCGPSPSPS 636
 QY 636 ANRFRITLLN-----LPAGVSPKHSPLIANTLWGNMLATESLKNKSA-----EL 684
 DB 637 AN---TLVAWNSDPTLVRSTYILDPERYGELVNSLW-----ISFLGNOAFSDIIOV 686
 QY 685 TPSPHFGITGGGLGMVYODPRENHPGFMHRSYSGAGMIAGQTH--TFSLSKSGQTYT 742
 DB 687 LLIHPIGSLITAKALGAVENHPRGHGFSGRYGQYQALSMNTDHTTGLSGQLXG 746
 QY 743 KLNERYAKNNVSKNY--SCQEMLFSLQEGFLTKLVGLVSYGCHNHFFYQGENL-T 799
 DB 747 KTNAPYDSCSEQWYLLSPGQFPITVOKSEALISWKAAGYXSKNHNTTYLRDPAK 806
 QY 800 SOGTRFROTNGAVFEDLPKMPFGSTHILAP-----FLGALGITYSSLSTFTEVGA 850
 DB 807 SOGQHHNNSYVVLISAE--HFELMWCLITRLLAQAMDLSGFIASEFLGQWQSKTEFGD 863
 QY 851 YPRSFSTPLINLVPIYGVKSGFNNATOR--POAWTVELAYOPVLYROEPIATOLLASK 909
 DB 864 LORGS-RGKGVNLSPLGCSQWFTPRKAPSTLTIKAIYRDPYRVAIPHIVVSNQ 922
 QY 910 GIWFGSGSPSSRHANSYKISQOTPLSLWTLFLFOYHGFYSSTPCNY 956
 DB 923 ESTSISGANLRHGLFVQI-HDVYDLTEDTQAFNLVTFGKNGFTNH 968
 RESULT 25
 Q9RB63 PRELIMINARY: PRT: 978 AA.
 AC Q9RB63:
 DT 01-MAY-2000 (Tremblrel. 13. Created)
 DT 01-MAY-2000 (Tremblrel. 13. Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 13. Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN H FAMILY (POLYMORPHIC MEMBRANE
 DE PROTEIN H FAMILY).
 DE PMP 14 OR CP0298.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.

OX NCBI_TaxID=83558:
 RN SEQUENCE FROM N.A.
 RC STRAIN=J138:
 RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CHL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39:
 RX MEDLINE=20150255; PubMed=10684935.
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Uterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouli H., Craven B.,
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
 RT pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL; AP002546; BAA98661.1;
 DR EMBL; AE002191; AAF38155.1;
 DR TIGR; CP0298;
 DR INTERPRO: IPR000515;
 DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1;
 SO SEQUENCE 978 AA; 103654 MW; 6EE142999D3019 CRC64;
 Query Match 11.5%; Score 583; DB 2; Length 978;
 Best Local Similarity 26.1%; Pred. No. 1.8e-31;
 Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
 QY 41 SNKISLTGDTNHL--TNCYL-DNLRYIIAIILOKTPNEGAATYTDYLSFPDOKGIYFA 97
 DB 67 SSNLSLKGKSLFTFTSOQAPTNVALL-----SAAETLT-FKNF-----SSINFT 112
 QY 98 KNLTPESGCAIYASPNSTPYEIRDTIGPVIFENN---TCCRPTSSNPNAVNKIREG 153
 DB 113 GMSOTGLGCLY-GCKD-----IVFOSIKDLIFTRVAVSPASVTSATP-AITVTGTG 164
 QY 154 G-----AIIAONLXIIHNHDVYGFMKNFSGVYGGAIATANTFVVS--ENOSCFLEMDN 204
 DB 165 ASALOPTDSLTVEN--ISOSIKIFGFLNLF-----GSAISSSTAVAKFINNTATMSFSHN 218
 QY 205 ICIOTNTAGKCAIYAGTSNFSFNCDLFFINNACC-----AGCAIFS 248
 DB 219 F-----TSSGGVYIGGSLFENNNGCIIFTANCVSLSKGVTPSSGTALGSGAICI 273
 QY 249 P--ICSLTGNRCNIYFVNNRCFKVETASSEASDCAIKVTRLDVTGNRCGRIFESDNT 306
 DB 274 PTGTFFELKNNGCKCTFSYN-----GTPNDAGAIYAEI-CNIVNOCALLDSDNTA 322
 QY 307 KNYGAIYAPVYTLVDNCPTEYFINNANKGAIYI-----DGTNSKISADRHAI 358
 DB 323 ARNGAICAKVANIIOGRPIESRRRA-EKGAIFIGSVGDPKAKOTSTLTLLASEGDI 381
 QY 359 FENIIVTNTANGSTSTANPPRRNAITVASSGCEILGAGSSONLIFYDPI-----E 411
 DB 382 FCGNML-----NTKPGIRNAITVEAGGEIVLSAOGGSLRYDPTTHSLPTTS 430
 QY 412 VSNAGSVSFNKEADQTSVVFSGATVNSAD-FHQRNLQTKTPAPLTLNSGFLCIEDHAQ 470
 DB 431 PSNKDITINAN---GASGVVFTSKGLSTELLIPANTTTLLGTGVKIASGELKITDNAY 487
 QY 471 LTVNRF-TOTGGVSLGNGAVLSCYKNGAGNSASNAITLKHIGLNSLSKSGEIFLL 529
 DB 488 VNVLFATOGSGQLTLGSGGTGLATPTGAPAA--VDFITGLADPSPFLKRD----- 539
 QY 530 WVEPTNNSNTYADTAATFSLSDVLSLIDYGNSPYESTDLTHALSSQPMLSISEASDN 589
 DB 540 FVSASVANG--TKNVTLTGAL-----VLDE-----HDVTDLYDNVSLQSPVAIPAIY-- 584

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 18:41:45 ; Search time 34.43 Seconds
(without alignments)
538,440 Million cell updates/sec

Title: US-09-677-752-2

Perfect score: 5086

Sequence: 1 MKKAFFFLIGNSLGLARE.....GFTSSSTFCNVLNGELALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summariesDatabase : Issued patents, AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194	3.8	1612	1	US-08-169-927-2
2	177.5	3.5	1338	2	US-08-728-470-9
3	177.5	3.5	1599	2	US-08-617-697-9
4	177	3.5	1477	1	US-08-038-682-4
5	177	3.5	1477	1	US-08-302-832-4
6	177	3.5	1477	1	US-08-530-198-4
7	177	3.5	1477	2	US-08-469-880-4
8	177	3.5	1477	2	US-08-728-470-4
9	177	3.5	1477	2	US-08-617-697-4
10	176.5	3.5	1536	1	US-08-038-682-2
11	176.5	3.5	1536	1	US-08-302-832-2
12	176.5	3.5	1536	2	US-08-530-198-2
13	176.5	3.5	1536	2	US-08-469-880-2
14	176.5	3.5	1536	2	US-08-728-470-2
15	176.5	3.5	1536	2	US-08-617-697-2
16	174	3.4	1160	4	US-08-808-599A-24
17	171.5	3.4	1529	2	US-08-728-470-10
18	171.5	3.4	1600	2	US-08-617-697-10
19	154.5	3.0	1026	1	US-08-194-290-7
20	150	2.9	983	2	US-08-164-292B-26
21	150	2.9	983	3	US-08-845-923-26
22	150	2.9	983	3	US-08-815-927-26
23	146.5	2.9	1026	2	US-08-614-377A-7
24	146.5	2.9	1026	4	US-09-142-648B-7
25	146	2.8	1007	4	US-08-551-459-4
26	144	2.8	1912	1	US-08-409-995-4
27	144	2.8	1912	3	US-08-685-467-4

28	144	2.8	2353	4	US-09-377-155-33	Sequence 33, Appl
29	144	2.8	2353	4	US-08-913-942-4	Sequence 4, Appl
30	143	2.8	890	2	US-08-483-101-14	Sequence 14, Appl
31	139	2.7	1394	5	PCT-US95-10661A-2	Sequence 2, Appl
32	138	2.7	1026	2	US-08-542-003-6	Sequence 6, Appl
33	138	2.7	1026	2	US-08-322-760A-6	Sequence 6, Appl
34	133.5	2.6	820	1	US-08-291-896-2	Sequence 2, Appl
35	133.5	2.6	820	2	US-08-485-278-2	Sequence 2, Appl
36	133	2.6	792	3	US-08-433-522A-8	Sequence 8, Appl
37	133	2.6	792	3	US-09-135-166-8	Sequence 8, Appl
38	131	2.6	704	3	US-08-792-832A-2	Sequence 2, Appl
39	131	2.6	797	3	US-08-433-522A-2	Sequence 2, Appl
40	131	2.6	797	3	US-08-433-522A-4	Sequence 4, Appl
41	131	2.6	797	3	US-08-433-522A-6	Sequence 6, Appl
42	131	2.6	797	3	US-09-135-166-2	Sequence 2, Appl
43	131	2.6	797	3	US-09-135-166-4	Sequence 4, Appl
44	131	2.6	797	3	US-09-135-166-6	Sequence 6, Appl
45	129	2.5	666	2	US-08-737-716-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-169-927-2
Sequence 2, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carlson, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2
Query Match 3.8%; Score 194; DB 1; Length 1612;
Best Local Similarity 18.9%; Pred. No. 2.3e-07;

QY 332 IANNKGA--IYIDGTSNKSISADRHAIIFENIYV-----NTVNAAGT-----ST 375
Db 807 ASGNADAKKVTFFKVKXKSDISTDGHNVTLNSEVKTSGSSNAGNDNSTGLTISAKDVT 866
QY 376 SANPRRRAIVTAVSSGELLGAGSSONLIFYDPIEVSNAGVSFENFADOTGSVFSG 435
Db 867 NNNVTSKTTINISAAAGVNTTKEGT-----INATGSEVY-AONGTIKGNITSON 917
QY 436 ATVSADFHQRLQKTPAPPLTSLNGFLCIEDHAQLTVNRFOTGQVSVLNGAVLSCYK 495
Db 918 VTYATE---NLVTENAVINATSG-----TVNISKTDI-----K 951
QY 496 NGAGNSASNAITLKHILNLSILKSGAEIPL-----WVEPTNNSNNYAD 543
Db 952 GGIESTSGNVNITASGNILKVSNI--TGODVTVTADAGALTITAGSTISATIGNNITTK 1009
QY 544 T-----AATFSLDVKLSLIDYGNS---PYESTDLTHALSSQPLMSISEASD 588
Db 1010 TGDINGKVESSGSVTLVATGATLAVGNISGNTVTTADSGKLTSTVGSTINGTNSVTT 1069
QY 589 NQLSDDMDFG--LNVPHYGOGLTWGMKTODPEPASATITDPOKANRFRHTL-- 644
Db 1070 SOSGIEGTISGNTVNTV-----ASTGLTIGNSAKV---EAKNGAATLTAE 1113
QY 645 ---LTWLPAGVVPSPKHSPLIA--NTLWGNMLATESLKNSAEITPSDHPWGTITGGL 699
Db 1114 SGKLTITQGSSTSSNGQTTLAKDSIAGINANAVLNTITGLTTGDSKINATSGTL 1173

RESULT 3
US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barekcamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Beikstreeser, Jeffery W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
US-08-617-697-9
Query Match 3.5%; Score 177.5; DB 2; Length 1599;
Best Local Similarity 21.4%; Pred. No. 6e-06;
Matches 154; Conservative 93; Mismatches 266; Indels 187; Gaps 33;
QY 75 GAATVITDYLSEFDQKEGIYFAKNLTPESSGALGASPNSEIYRDTGPVIFENNTC 134
Db 806 GGNITIFQVEGTDSRVNKGVAKKNIT-FKGNITFGGOKA-TTEIK---GNVTINKNTN 860
QY 135 CRFTSSNPAVNR-----IRCGAI-HAQNLYT-----NHHMD 168
Db 861 A---TLRGANFENKSPINAGNVINNGLTAGSIINAGNLTVSKGANLQAITNTYFN 917
QY 169 VYGFV-----KFEYVVRGA-----ISTANTFVSENOGCFLEMDICIoTNTAGKGA1Y 219
Db 918 VAGSEFNNCASNIS1ARGAKAFKIDINNTSSLNITNSDTTY---RT1IKGNISNKG--- 971
QY 220 AGTSNSESNNCDLFFINNACCA---GGAIFSPICSLTGNKGNIVFYNNRCFK-NVE-- 272
Db 972 -----DLNIDKSDAEIQIGNISQKEGNLTISSDKVNITNQTITKAGVEGG 1019
QY 273 -TASSEASDGCALKYTRLDVTGNKGRIFPSDNITKNYGALYAPVTVLVDNGPTFTNN 331
Db 1020 RSDSSEAEANLTIOTKELKAGD-----LNISGFKAET--AKNGSDLTIGN 1066
QY 332 IANNKGA--IYIDGTSNKSISADRHAIIFENIYV-----NTVNAAGT-----ST 375
Db 1067 ASGNADAKKVTFFKVKXKSDISTDGHNVTLNSEVKTSGSSNAGNDNSTGLTISAKDVT 1126
QY 376 SANPRRRAIVTAVSSGELLGAGSSONLIFYDPIEVSNAGVSFENFADOTGSVFSG 435
Db 1127 NNNVTSKTTINISAAAGVNTTKEGT-----INATGSEVY-AONGTIKGNITSON 1177
QY 436 ATVSADFHQRLQKTPAPPLTSLNGFLCIEDHAQLTVNRFOTGQVSVLNGAVLSCYK 495
Db 1178 VTYATE---NLVTENAVINATSG-----TVNISKTDI-----K 1211
QY 496 NGAGNSASNAITLKHILNLSILKSGAEIPL-----WVEPTNNSNNYAD 543
Db 1212 GGIESTSGNVNITASGNILKVSNI--TGODVTVTADAGALTITAGSTISATIGNNITTK 1269
QY 544 T-----AATFSLDVKLSLIDYGNS---PYESTDLTHALSSQPLMSISEASD 588
Db 1270 TGDINGKVESSGSVTLVATGATLAVGNISGNTVTTADSGKLTSTVGSTINGTNSVTT 1329
QY 589 NQLSDDMDFG--LNVPHYGOGLTWGMKTODPEPASATITDPOKANRFRHTL-- 644
Db 1330 SOSGIEGTISGNTVNTV-----ASTGLTIGNSAKV---EAKNGAATLTAE 1373
QY 645 ---LTWLPAGVVPSPKHSPLIA--NTLWGNMLATESLKNSAEITPSDHPWGTITGGL 699
Db 1374 SGKLTITQGSSTSSNGQTTLAKDSIAGINANAVLNTITGLTTGDSKINATSGTL 1433

RESULT 4
US-08-038-682-4
Sequence 4, Application US/08038682
Patent No. 5549897
GENERAL INFORMATION:
APPLICANT: BAREKAMP, STEPHEN J
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia

```

Db 648 TRKNTSYWQTHSDHWN/SALNLETGANFTFKIYSSNSKGLTTOYRSSAGV----- 699
QY 135 CRPTSSNPNAVNKIREGAIHAO-----NLXI 163
Db 700 --NFGVNGMNSFN-LKEGAKVNFLEKPNENNTSKPLPIRFLANITATGGSVFFDIYA 756
QY 164 NHN-----HDVGFPMKNSFYRGAT-----STANTFVVSNOQCL 200
Db 757 NHSGGAEIKMSEINISGANTFLN-SHYRGDAFKINKDLTINATNSNFSLRQTKDDEFY 815
QY 201 --FMDNICIQT-NTAGKGAIIYAGTSNFSNCDLFF-----INNA----- 239
Db 816 DGYARNAINSTINISILGCVNTLGGONSSSITGNITTEKAAVNTLEANNAPNOQINRDR 875
QY 240 -CCAGGAIFSPICSLTGN-----RCNIVFYNNRCFK-----NVEETASSEASD 280
Db 876 VIKGLSLVNGSLSLTGENADIKGNLTISESATFKGKTDTINTGNFTNGTAELINTO 935
QY 281 G-----GAIKVTTRL-----DVTGNRGRIFFSD----- 303
Db 936 GVVKLGNTNDGDLNITTHAKRNORSIIIGDIIINKGSLNITDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGGAIIYAPVTLVD----- 322
Db 996 GNLTISSDKINITKOITIKKIDGEDSSDATSNANLTITKTELKLTEDLSISGFNKAEI 1055
QY 323 --NGPTYFINNIANNKGA-----IYIDGTSNKSISADRHAIIFENIYVNTYNNANGTST 375
Db 1056 TAKDGRDLTIGNSNGNSGALEKATYFNNVKDSKISADCHNTLNSKYVTSSNGRGREN 1115
QY 376 SANPRRANITVA-----SSGEILLGAGSSONLIYDPIEVSNAGV 417
Db 1116 SDND---TGLITTAKNVEYNKDITSLKTIVNITASEKVTTTAGST-----INATNGKA 1164
QY 418 SVSFNKADQTSVVFSGATVN-SADFHORNIOTKTPAPLTLISNGFLCIEDHAQL----- 471
Db 1165 SIT-TKTGDISGTI--SGNTVSATV---DLTTSGSKIEKSGEANTVSATGTIGCTI 1218
QY 472 --TVNRFOTOGVYSLGCAVILSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPL 528
Db 1219 SCNTYV-NVANGDGLTVGCAEINATEGATLTLATGNLTITE-AG---SISITKQOVL 1273
QY 529 L---WEPTNNSNNTADTAATFSL---SDVKL---SLID-----IGNSPYESTDLT 572
Db 1274 LAONGSIAGSIANAANTLTMTGTLTVAOSDIKATSGTLVINAKDAKLMDASCDSTEV- 1332
QY 573 HALLSOPMLISEASON--QLRSDMDPFGSLAV 603
Db 1333 NANNASGSGSVTAATSSVNTIGDLNTVNGLNI 1365

RESULT 6
US-08-530-198-4
Sequence 4, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

Query Match 3.5%; Score 177; DB 2; Length 1477;
Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

QY 37 KESLSNKLISLTD-----THNLTCYLDNLRLYLALQKTPNEGAATITD- 82
Db 603 KOFRRANNVSLNLTGKGLNLISSVNNLTHLSG-----TINISGNTINTOT 647
QY 83 ---YLSFDTQKEGIIYFAKNLTPESSG---AIGVSPNSP--TVEIRDTIGYIENNCT 134
Db 648 TRKNTSYWQTHSDHWN/SALNLETGANFTFKIYSSNSKGLTTOYRSSAGV----- 699
QY 135 CRPTSSNPNAVNKIREGAIHAO-----NLXI 163
Db 700 --NFGVNGMNSFN-LKEGAKVNFLEKPNENNTSKPLPIRFLANITATGGSVFFDIYA 756
QY 164 NHN-----HDVGFPMKNSFYRGCAI-----STANTFVVSNOQCL 200
Db 757 NHSGGAEIKMSEINISGANTFLN-SHYRGDAFKINKDLTINATNSNFSLRQTKDDEFY 815
QY 201 --FMDNICIQT-NTAGKGAIIYAGTSNFSNCDLFF-----INNA----- 239
Db 816 DGYARNAINSTINISILGCVNTLGGONSSSITGNITTEKAAVNTLEANNAPNOQINRDR 875
QY 240 -CCAGGAIFSPICSLTGN-----RCNIVFYNNRCFK-----NVEETASSEASD 280
Db 876 VIKGLSLVNGSLSLTGENADIKGNLTISESATFKGKTDTINTGNFTNGTAELINTO 935
QY 281 G-----GAIKVTTRL-----DVTGNRGRIFFSD----- 303
Db 936 GVVKLGNTNDGDLNITTHAKRNORSIIIGDIIINKGSLNITDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGGAIIYAPVTLVD----- 322
Db 996 GNLTISSDKINITKOITIKKIDGEDSSDATSNANLTITKTELKLTEDLSISGFNKAEI 1055
QY 323 --NGPTYFINNIANNKGA-----IYIDGTSNKSISADRHAIIFENIYVNTYNNANGTST 375
Db 1056 TAKDGRDLTIGNSNGNSGALEKATYFNNVKDSKISADCHNTLNSKYVTSSNGRGREN 1115
QY 376 SANPRRANITVA-----SSGEILLGAGSSONLIYDPIEVSNAGV 417
Db 1116 SDND---TGLITTAKNVEYNKDITSLKTIVNITASEKVTTTAGST-----INATNGKA 1164
QY 418 SVSFNKADQTSVVFSGATVN-SADFHORNIOTKTPAPLTLISNGFLCIEDHAQL----- 471
Db 1165 SIT-TKTGDISGTI--SGNTVSATV---DLTTSGSKIEKSGEANTVSATGTIGCTI 1218
QY 472 --TVNRFOTOGVYSLGCAVILSCYKNGAGNSASNASITLKHIGLNLSSILKSGAEIPL 528

```

APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

Query Match 3.5% Score 177; DB 2; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

QY 37 KESLSNKLSTLGD-----THNLNCLYLDNLRVILAILOKTPNEGAATITD- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTHNLSC-----TINISGNTITNOT 647
QY 83 ---YLSFDTQKEGIFYFAKNLTPESG---AIGVAPNSP--TVEIRDTIGPVIFENNTC 134
DB 648 TRKNTSYWOTSHDSHMVNSALNLETGANFTIKYISSNSKGLTQYRSSAGV----- 699
QY 135 CRPTSSPNAVAVKIRREGCAIHAQ-----NLXI 163
DB 700 --NENGVAHNSFN-LKEGAVNEFLKPNENMNTSKPLRIFLANITATGGGVFEFDIYA 756
QY 164 NNN-----HNVGFMKNFSYVRCGAI-----STANTFVVENOSCFI, 200
DB 757 NMSGRAELKSEINISNCAFTLN-SHVRGDDAKINKDLTINATNSFSLRQKDFY 815
QY 201 --FMDNICIQ-NTAGKGAIVAGTSNFSFNCDLF-----INNA----- 239
DB 816 DGYARNAINSTYINISILGCVTLTGONSSSITGNITTEKAAVTLLEANNAPNOQNIHQR 875
QY 240 -CCAGAGAFSPICSLTGN-----RGNIFYNNRCEK-----NVTASEASD 280
DB 876 VIKUGSLLVNGLSLTGENADIKGLNLTSESATFEKGRDRLNLTGNFTNGTAIINITQ 935
QY 281 G-----GAIVTTRL-----DVTGRCRIFFSD----- 303
DB 936 GVVKLGANTNGDNLITTHAKRNRQSIIGDILINKKGSINLIDSNDAEIQIGNISQKE 995
QY 304 -----NITK-----NYGAIYAPVTLVD----- 322
DB 996 CNLTISDKINITKOITIKKIDGEDSSSDATSNANLTIKTEKLITLEDLSISGFENKAEI 1055
QY 333 ---NCPYFINNIANNKGA---TYIDGTSNKSISADRHAIIFENIYVNTNANGST 375
DB 1056 TAKDREDLTLGNSDNGSGAEAKIVTFPNVYKDSKISADGHVTLNLSKRYKTSSNGGRESN 1115
QY 376 SANPRRRAITVA-----SSGEILLGAGSSONLIFYDPIEVESNAGV 417
DB 1116 GUND---TGLITTAKNVEVKNKDIISLKTIVNITASEKVTJTAST-----INATNCA, 1164
QY 418 SVSFNKEADQIGSVFSGATVA-SADFHQRNLQTKTPAPRLTSLNGFLCIEDHAOL----- 471

DB 1165 SIT-TKTDIGISGII--SGNTVSATV---DLTKSGSKIEAKGSEANVSATGTIGCTI 1218
QY 472 ---TVNRFYQGVVSLGNGAVLSCKYKNGAGNSASNAITLKHIGLNLISLKSAGAEIPL 528
DB 1219 SGNTVN-VTANAGDLTYGNGANEINATEGAATLTATGNTLTTE-AG---SITSTKGGVDL 1273
QY 529 L---WPEPTNSNNYIADTAATFSL---SDVKL---SLDD-----YGNSPRESTDLT 572
DB 1274 LAONGSIAGSINAANVTLTGTLTVAAGSDIKATSGTLVINAKDARLNGASGDSSTEV- 1312
QY 573 HALSSQPLSLISEASDN--QLRSDMDSESGNV 603
DB 1333 NAVNAGSGSVTAATSSSVNITGDLNVTNGLNI 1365

RESULT 9

US-08-617-697-4
Sequence 4, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: Of No. 5977336-Typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstesser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-4

Query Match 3.5% Score 177; DB 2; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

QY 37 KESLSNKLSTLGD-----THNLNCLYLDNLRVILAILOKTPNEGAATITD- 82
DB 603 KDFRANNVSLNCTGKGLNISSVNNLTHNLSC-----TINISGNTITNOT 647
QY 83 ---YLSFDTQKEGIFYFAKNLTPESG---AIGVAPNSP--TVEIRDTIGPVIFENNTC 134

; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ. ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-10

Query Match 3.4%; Score 171.5; DB 2; Length 1529;
Best Local Similarity 23.5%; Pred. No. 1.8e-05;
Matches 104; Conservative 70; Mismatches 186; Indels 83; Gaps 21;
QY 139 TSSNPNAAVKIREGCAIHAQNIYINHN---HDVYGFKNFSYRGCAISTANTFVYSEN 195
DB 1033 TSNSSNAGNDNSTGLTISAKDVTNNNTVSHKTI---MISAAGNVTTKEGTTINATT 1088
QY 196 QSCFLEFNDNICIQTNTAGKGAIIAGTSNFSFESNCDLFFIN-----NCCAGCAIFSP 249
DB 1089 GSVEVTAQNTIGNTISQNVYTA--TENLVTTENA---VINATSGTVNISTKTGDIKGC 1144
QY 250 ICSLUTGN----RGNIFYNNRCFKNVTASSSEASDGAIKVTRLDVTGNRGRIFESDN 304
DB 1145 IESTSGNVNTASGNTIKVSNITGQV---TYADAGALTITTAGSTISATG---NAN 1196
QY 305 ITKNYCG---GAIYAP--VTLVDNGPTFTYNNINANNKGAIIYDGTSSNSKISADRAHAIIF 359
DB 1197 ITTKTGINKVESSESGSVTLVATGATLAVGNISGN-----TVTITADSGKLTSS 1245
QY 360 NENIVTNTANGTSTANPR-----RNATIVASSSGEILLGAGSSONLIFYDIEVSN- 414
DB 1246 TVGSTINGTNSVTSSQSGDIEGTISGNIVNVITASTGDTLTIGNSAK-----VEAKNG 1297

OY 415 -AGVSVSFNKEADOTGSVVS--GATVNSA-----DFHQRNIQTKTPAPLTLNSGFLC 464
DB 1298 AATLTAESEKLTOTGSSITSSNGOFTLLTAKDSSIAGNINAAVNTLTUTGTLTGDSKI 1357
QY 465 IEDHAQTVN-RFTQNGCVVSLNGAVLSCYKNGAGN---SASMSIT--LKHI-GLNL 516
DB 1358 NATSGTLTINAKAKLDGASGDRYVATNAGSGNVTAKTSTSSVNTIGDLTINGLNI 1417
QY 517 -----SILKSGAEIPLWVEP 533
DB 1418 ISENGRNTVRLRKEIDVKYIQP 1440

RESULT 18
US-08-617-697-10
; Sequence 10, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ. ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1600 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-617-697-10
Query Match 3.4%; Score 171.5; DB 2; Length 1600;
Best Local Similarity 23.5%; Pred. No. 2e-05;
Matches 104; Conservative 70; Mismatches 186; Indels 83; Gaps 21;
QY 139 TSSNPNAAVKIREGCAIHAQNIYINHN---HDVYGFKNFSYRGCAISTANTFVYSEN 195
DB 1103 TSNSSNAGNDNSTGLTISAKDVTNNNTVSHKTI---MISAAGNVTTKEGTTINATT 1158
QY 196 QSCFLEFNDNICIQTNTAGKGAIIAGTSNFSFESNCDLFFIN-----NCCAGCAIFSP 249
DB 1159 GSVEVTAQNTIGNTISQNVYTA--TENLVTTENA---VINATSGTVNISTKTGDIKGC 1214

TELEX: 34-0154
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 983 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-164-292B-26

Query Match 2.9%; Score 150; DB 2; Length 983;
 Best Local Similarity 19.8%; Pred. No. 0.00061;
 Matches 199; Conservative 108; Mismatches 392; Indels 304; Gaps 46;

```

188 NTFVSENOGCEFLPMONICQITNTACKGAIYAGTSNSESNNCDLFFINMCCAGGAIF 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 NGFV--ENGEATLAM--LVEKPLTFDEKAGALTGVGRGIRINPAGLETND--LASAVF 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 SPICSLTGKRGIVVFYNNCFKVNFTASSSEASDGAIKVTRTRDVTGGR-----GRI 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 PELAS--DEAGVNTL-----NMSDGLYTKDKKLAVKVGPGSLDSNNALQVHTGDGLT 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 FESDNTKKNYGAIVAVYT-----LVNDGPTYFINN--IANKK 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 VTDDKVS LN-----TQAPLSTTSAGLSLLGPSLHGEERLTVNTGAGLQISNNALAVKY 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 GGAIIYDGTNSKIS--ADRHAIIFENIY-----TN--VTNANGTSTANPPRRNAI 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 GSGITVDAQNOQLAASLGDLSESRDNTVYKAGPGLTITNOALTVAATGNGLOVNEGOLQ 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 TVASSSG-----ELLGAG-----SSONLIIFYDPIEVSNAGVSFKNKEADQTS--- 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 NITAGGGLNFANNSLAVEGSLGFPPGQNOVSLY-----GDCIDIRDKRVVTPAGPGLR 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 -----VFSGATVNSADFHQRNLQTKTPAPLTLSNGF-----LCIEDHAQLTVNRF 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 MLNHOALVAVSG--DGLEVHSDTLRLKLSHGLTFENGAVRAKIGPGLGIDDSGRSVY--- 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 TOTGGVAVSLGNGAV--LSCYKKNAGNSAS--NASITLKHIGLNLSSILKSGAEI----- 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 -RTGRGLRVAANGVOIFSGRGTALIGTDSLTLLIRAPLQFSGPALTAASLGSGPITYNSN 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 -----PLMWEPNT-----NSNNYTDATATFSLSDYVKLSLIDYGNSPY 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 NGTFGLSIGPGMWVDNRLOVNPAGLIVFOGNNLVNPLADPLAISDSKISLSLGPGLT-Q 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 ESTDLTHALS-----SOPMLSTIS-----EASDNLRSDDMDFSGLVN-----PHYG- 607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
485 ASNALTLISLGNGLFESNOVAIKAGRGLREFSSQALLESSLTIVGNGLTITLDVYIRPNLGD 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
608 -----WQGLMTWGMAKTODPEPASSATITDPOKANEHRTL----- 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
545 GLEVRDNKIIVKLGANLRENGAVTAGTVPSAPEAPRTITAEPRILASHSLOLISEG 604
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
644 -----LTMPLAGYVPSIRKHSRPL----- 662
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
605 LVVHNNALALQJGDGMEVNOHGILTLRVGSGLOQRDGLITVTPSGTRLEPRILAPLOTEN 664
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
663 -IANTLGMMLLATESLK-----NSAELTPSDHFWGTLGGGLGMAYUODREHNPFGHMR 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
665 GIGLALGAGLEDESALQYKVGPGMRLNPEKVVTLILGPGLS--FGQD-AARTNVNDR 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
718 SSGYSAGMAGOT-----HTFSLKFSQTYTKLNERVAKNNVSSKNSCOGEMLEFSLQ 769
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
721 VS--VEPPNVFGGQGLTFLVGHGLIIONSRLQNLGGLRTDPTVQLQLEVPLOQGLEIAD 779
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
770 EGFLLTKLVGLVSYGD--HNCHEFTQGENLTSSOG-----TFRSOTMGCAVVF 815
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780 ESQVRAKL-----GGLQLOFDSQARTITAPNMVTELTMTGTGSNANVTMGYTAAPGSKLF 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
816 DLPMKPFGSTHILTAFLGALGIYSSLSHTEFGAVPRSESTKTPILNVLVPIGVKGSFM 875
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 -----LSLTRYSTGLVIGMTIDSNAS-----FGQY--INACHFQJECFTLLDNGCNLK 880
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

876 NATQRPQAMTVE-----LAPQVLYRQE---PGIATOLLASKGIMFGSG---- 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 881 EGSNLOGTWEKKNPNPASKAAFLPSTALYPLINESGSLPG--KNLVGMQAILGGGGTCT 938
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
917 ----SPSSRHAMSYSKISQO-----TOPLSWLTLPQY 944
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 939 VIATLNGRRSNNYPACOSIIFVWQEFNTIAROPLANNSTLTFFSI 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 21
 US-08-845-623-26
 Sequence 26, Application US/08845623A
 Patent No. 6001591

```

GENERAL INFORMATION:
APPLICANT: BABIRK, LORNE A.
APPLICANT: TIKOO, SURESH K.
APPLICANT: REDDY, POLICE S.
TITLE OF INVENTION: BOVINE ADENOVIRUS 3 GENOME
FILE REFERENCE: 293102002120
CURRENT APPLICATION NUMBER: US/08/845,623A
CURRENT FILING DATE: 1997-04-25
EARLIER APPLICATION NUMBER: 08/164,294
EARLIER FILING DATE: 1993-12-09
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 983
TYPE: PR
ORGANISM: Bovine adenovirus type 3
US-08-845-623-26

```

Query Match 2.9%; Score 150; DB 3; Length 983;
 Best Local Similarity 19.8%; Pred. No. 0.00061;
 Matches 199; Conservative 108; Mismatches 392; Indels 304; Gaps 46;

```

188 NTFVSENOGCEFLPMONICQITNTACKGAIYAGTSNSESNNCDLFFINMCCAGGAIF 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 NGFV--ENGEATLAM--LVEKPLTFDEKAGALTGVGRGIRINPAGLETND--LASAVF 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 SPICSLTGKRGIVVFYNNCFKVNFTASSSEASDGAIKVTRTRDVTGGR-----GRI 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 PELAS--DEAGVNTL-----NMSDGLYTKDKKLAVKVGPGSLDSNNALQVHTGDGLT 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
300 FESDNTKKNYGAIVAVYT-----LVNDGPTYFINN--IANKK 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
141 VTDDKVS LN-----TQAPLSTTSAGLSLLGPSLHGEERLTVNTGAGLQISNNALAVKY 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
337 GGAIIYDGTNSKIS--ADRHAIIFENIY-----TN--VTNANGTSTANPPRRNAI 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 GSGITVDAQNOQLAASLGDLSESRDNTVYKAGPGLTITNOALTVAATGNGLOVNEGOLQ 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 TVASSSG-----ELLGAG-----SSONLIIFYDPIEVSNAGVSFKNKEADQTS--- 430
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 NITAGGGLNFANNSLAVEGSLGFPPGQNOVSLY-----GDCIDIRDKRVVTPAGPGLR 312
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 -----VFSGATVNSADFHQRNLQTKTPAPLTLSNGF-----LCIEDHAQLTVNRF 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
313 MLNHOALVAVSG--DGLEVHSDTLRLKLSHGLTFENGAVRAKIGPGLGIDDSGRSVY--- 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
477 TOTGGVAVSLGNGAV--LSCYKKNAGNSAS--NASITLKHIGLNLSSILKSGAEI----- 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
367 -RTGRGLRVAANGVOIFSGRGTALIGTDSLTLLIRAPLQFSGPALTAASLGSGPITYNSN 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 -----PLMWEPNT-----NSNNYTDATATFSLSDYVKLSLIDYGNSPY 566
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
426 NGTFGLSIGPGMWVDNRLOVNPAGLIVFOGNNLVNPLADPLAISDSKISLSLGPGLT-Q 484
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 ESTDLTHALS-----SOPMLSTIS-----EASDNLRSDDMDFSGLVN-----PHYG- 607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
485 ASNALTLISLGNGLFESNOVAIKAGRGLREFSSQALLESSLTIVGNGLTITLDVYIRPNLGD 544
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:09:45 ; Search time 63.9 Seconds
(without alignments)

1037.835 Million cell updates/sec

Title: US-09-677-752-2

Perfect score: 5086

Sequence: 1 MKKAFFFLIGNSLGLARE.....GFYSSTFCNLTNGELALRF 965

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4089.5	98.1	964	2	E71460
2	3648.5	71.7	976	2	F81722
3	1120.5	22.0	928	2	H72074
4	1060.5	20.9	952	2	D81593
5	1052	20.7	934	2	A72075
6	951.5	18.7	946	2	D81594
7	948.5	18.6	946	2	C72075
8	866.5	17.0	1034	2	F71460
9	793	15.6	1025	2	G81722
10	655.5	12.9	1013	2	G71460
11	645	12.7	936	2	C72078
12	644	12.7	936	2	B81591
13	620.5	12.2	987	2	H81722
14	620	12.2	930	2	A81591
15	617	12.1	930	2	D72078
16	597	11.7	928	2	D72077
17	597	11.7	949	2	F81591
18	595	11.7	1016	2	H71460
19	586.5	11.5	928	2	G81591
20	586	11.5	947	2	D72067
21	585.5	11.5	928	2	B72077
22	583	11.5	978	2	G72076
23	583	11.5	978	2	B81593
24	575	11.3	973	2	F72076
25	575	11.3	995	2	C81593
26	570.5	11.2	922	2	B72131
27	569.5	11.2	922	2	F81539
28	567	11.1	1276	2	C81591
29	564	11.1	983	2	A81723

30	562	11.0	1407	2	B72078	polymorphic outer
31	546.5	10.7	986	2	B81675	polymorphic membra
32	525.5	10.3	841	2	E72130	polymorphic membra
33	519.5	10.2	975	2	F71518	hypothetical prote
34	517	10.2	1723	2	E72067	polymorphic membra
35	517	10.2	1732	2	C81601	polymorphic membra
36	512	10.1	392	2	B72075	polymorphic outer
37	474.5	9.3	1609	2	H72013	polymorphic membra
38	470.5	9.3	878	2	B71460	probable outer mem
39	457	9.0	867	2	F81721	polymorphic membra
40	395	7.8	1751	2	G71518	hypothetical prote
41	389.5	7.7	1672	2	C81675	polymorphic membra
42	386	7.6	1531	2	H71468	probable outer mem
43	377.5	7.4	1460	2	D81675	polymorphic membra
44	358.5	7.0	1520	2	A81731	polymorphic membra
45	342	6.7	1770	2	A71517	hypothetical prote

ALIGNMENTS

RESULT 1
E71460
probable outer membrane protein E - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence-revision 13-Sep-1998 #text-change 08-Oct-1999
C:Accession: E71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MIMID:99000809
A:Accession: E71460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-964 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:93329342; PIDN:AA068467.1; PID:9332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmpe

Query Match 98.1% Score 4989.5; DB 2; Length 964;
Best Local Similarity 98.3% Pred. No. 9.2e-299;
Matches 949; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

QY	1	MKKAFFFLIGNSLGLAREVPSRIPLMPNSVDPPTKESLSNKISLTGDTNHLNLCYLDN	60
DB	1	MKKAFFFLIGNSLGLAREVPSRIPLMPNSVDPPTKESLSNKISLTGDTNHLNLCYLDN	60
QY	61	LRYLAILOKTPNGCAVTTIDYLSFFDQKEGYIFAKNLTPEGGAIGYASPSPTVEI	120
DB	61	LRYLAILOKTPNGCAVTTIDYLSFFDQKEGYIFAKNLTPEGGAIGYASPSPTVEI	120
QY	121	RDITGPVIEFNNTCCRPETSSNPNAVANKIREGAIHONLIYHNHNDVGFPMKFSYVR	180
DB	121	RDITGPVIEFNNTCCRPETSSNPNAVANKIREGAIHONLIYHNHNDVGFPMKFSYVR	180
QY	181	GCAISTANTFVVSNSQCFLEMDNICIOTNTAGKGAIVYACTSNSPESNCDLFFINNAC	240
DB	181	GCAISTANTFVVSNSQCFLEMDNICIOTNTAGKGAIVYACTSNSPESNCDLFFINNAC	240
QY	241	CAGGAIFSPICSLTGNRNIIVFYNNRCKNKTETASSESDGCAIKVTRLDVNTNRCRIF	300
DB	241	CAGGAIFSPICSLTGNRNIIVFYNNRCKNKTETASSESDGCAIKVTRLDVNTNRCRIF	300
QY	301	FSDNTKNGAIIAPVYTVLDNGPTTFINNANKKGAIIYDGTSNKISADRAHAIIFN	360
DB	301	FSDNTKNGAIIAPVYTVLDNGPTTFINNANKKGAIIYDGTSNKISADRAHAIIFN	360
QY	361	ENIVTNTNANGTSTANPPRRNAITVVAASGGEILLGAGSSQNLIFYPDIEVSNAQSVS	420
DB	361	ENIVTNTNANGTSTANPPRRNAITVVAASGGEILLGAGSSQNLIFYPDIEVSNAQSVS	420
QY	420	ENIVTNTNANGTSTANPPRRNAITVVAASGGEILLGAGSSQNLIFYPDIEVSNAQSVS	480
DB	420	ENIVTNTNANGTSTANPPRRNAITVVAASGGEILLGAGSSQNLIFYPDIEVSNAQSVS	480

A:Molecule type: DNA
 A:Residues: 1-938 <ARN>
 A:Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PID:AA018608.1; PID:9437675
 A:Experimental source: strain CW1029
 R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; M0ID:20150255
 A:Accession: E81593
 A:Molecule type: DNA
 A:Residues: 1-938 <REA>
 A:Cross-references: GB:AE002190; GB:AE002161; NID:97189209; PID:AAF38143.1; PID:9718921
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: pmp_15; CP0286

Query Match 22.0%; Score 1120.5; DB 2; Length 938;
 Best Local Similarity 30.9%; Pred. No. 5e-61;
 Matches 308; Conservative 163; Mismatches 418; Indels 109; Gaps 31;

6 FFFLLGSLSLGAREVSRIFELMNSVPDPTKESLSKISLTGDTHTLT-----VGFTHNQNDLAIYGNHD 59
 12 FTVVLAN--EGLQLPLETYITLSPYQAAQ-----NC 56
 57 YLDLRYLILLOKTPNEGAAYTITDYLSFEDTOKEGYFAKNLTPESGAGIYASPNP 116
 60 FILDYKY-----RSNGALGCKNL--ISENIGVFEKKVCPNSGAI-YAAQNC- 108
 117 TVEIRDTIGVIEENNTCCRPFTSSNNAVANKIREGAIHAONLNIHNDVYFKNF 176
 109 -----TISK--NQNYAFTNLVSDNPTAGSL--GALFAIKCSTNNMGCGTFVNDL 159
 177 SYVKGAIKSTANFVSENOCSFLFMDNICIOTNTAGKGAIVAGTNSFESNCDLFEI 236
 160 ALNKGALYETNLSIDNKGPIIKONRAL--NSDSIGGCIYSGNSLIEGNSGAIQT 217
 237 NNACGAGATFS-PICSLTGNRGNIVFYNNCFKNVETASSEASDGAIKYTRLDVTGN 295
 218 SNSGSGGIFSTQTLTSSNKKLIEISENCAFAN--NYGSNPNPGGGLTTTCTILNN 275
 296 RGRFFESDNTKNGGAIYAPVTLVNGPTFYNNIANNKGAIYD-GTSNKK--ISA 352
 276 REGVLFNNSQOSNGAIHAKSIIEKNGPYFLNNTAIGGALLNSAGSGNSFLISA 335
 353 DRHAIIFENIVTNTANGSTSANPRRAITVASSSGEILLAGASSQNLFIYDPIEV 412
 336 DNGDIFRNN-----TASKHALNPYRNAIH-STPNMMLQIGARGYVLETPDIH 386
 413 S-NGVSVFENKADOTGVSFSGATVNSADFHORNIOTKTPAPLTLNNGFLIEDHAOL 471
 387 ELRSPFLIFNEFEGHGTVLFEGSEHVHONFTDEMNFSYLNRNSELNGVLAVEDGAGL 446
 472 TVNFTOTGGVSLGNCVAVSLCYKNAGN-----SASNAITLKATIGLNSILKSG 523
 447 ACYFEGFGGTLTGOGAVITTT--AGTIPSPSTPTVSTLTLNIAIDLSLISFQ 502
 524 AEIILLWEPTNNSNNTADPAATFSLSDVLSLIDDGNSPYESTDTHLSSOPMLSI 583
 503 AOAIRKIMYPTKGTSTYEDSNPTITISGT-LTLRNSMNEPDBDLSHLEKVPFLYI 561
 584 SEASNDLRSDMDPFSGLNV-PHYGMOGLWTGWAKTODPEASSATITDPQK--ANRF 639
 562 VDVAQAOKINSQDLSTLNSGEHYGQIGSTYVET-----TTITNPTSLGANTK 613
 640 HRTLLTLTPAGIYVSPKRRSPLANTLGNMLATBESLKSASALTLPDHPWGTITGG-- 697
 614 HKLLYANNSPLGYRHPRREREFITNALMQSAYTLAAGLSLS-----WDEEKSHA 665
 698 -----GLGMVYVODPRENHPGFHMRSSGSAGMIA--GOTHFSLKFSOTYTLNRYAKN 751
 666 ASLOGIGLLVHOKKNGKGRFRSHMTGYSATTEATISSQSPNLSGFAQFSAKAKHESON 725

752 NVSSKNV-----SCGEMLFSLQEGFLLTKLVGL-YSYGDHNCCHFFYTOGENTLSOGERSO 807
 726 STSHHHYFGMCIENTLE--KENIRLS--VSLAYMTSEHTHTMYGGLGNSQGSFHHN 781
 808 TMGCAVFFDPMKPRFSTHILTPAFGLAGIYSSLSHTEVGAYPSPSFSTKPLIANVLP 867
 782 TLGALSCVFLPQPHES-LQIYFPIITALAIRGNLAFOGSDHAFESILMRPLTIVSLP 840
 868 IGVGSGFMATQRPQAMVELAYOPVLYRQEPGIAFOLLASKGIMFGSGSPSRHAMSKT 927
 841 VGIASMKNNHRYPLVWLTETISYRSTLYRQDPRLSHKSLISQGTWTQATPYNALGK 900
 928 ISOOTQPLSLTLHPFOYHNGFYSSSTCNLYNGELALRF 965
 901 VKNTMOVFPKVTLSLDYSADISSSTLSHLYNVASRMKF 938

RESULT 4
 DB1593
 polymorphic membrane protein E/F family CP0286 [imported] - Chlamydia pneumoniae
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 11-May-2000
 C:Accession: DB1593
 R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: AB1500; M0ID:20150255
 A:Accession: DB1593
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-952 <REA>
 A:Cross-references: GB:AE002190; GB:AE002161; NID:97189209; PID:AAF38142.1; PID:9718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0285

Query Match 20.9%; Score 1060.5; DB 2; Length 952;
 Best Local Similarity 30.6%; Pred. No. 2.5e-57;
 Matches 303; Conservative 164; Mismatches 447; Indels 77; Gaps 28;

6 FFFLLGSLSLGAREVSRIFELMNSVPD-----PTKESLSK-----ISL 46
 8 FLYLGNFTACMGMTPAVYSLQDSLEKFALEDEEFRTSFLDLSLTLTGFSPTTF 67
 47 TGDTHNLT-NCYLDNRLYILLOKTPNEGAAYTITDYLSFEDTOKEGYFAKNLTPESG 105
 68 VGNRRHNSQDVLNSNYSIDNILLMTSAGAVACNNFL--LSNVEDHAFPSKALGTC 125
 106 GAIGYASPNSTPVEIRDTIGVIEENNTCCRPFTSSNNAVANKIREGAIHAQ-NLYIN 164
 126 GAIACQG---ACTITTKRGPILFFSNRGLN-----NASTGETRGCAIACNGDFTS 174
 165 HNHVYGFMKFSTYRGCAISTANTFVSENOCSFLFMDNICIOTNTAGKGAIIYAGTSN 224
 175 ONOGTEFFVNNSVNNMGALSTNGHCRIOSNRAPLFEFN-----NTAPSGGALRENT 228
 225 SFESNCDLFEINNACGAGATFSPI-CLSLTGNRGNIVFYNNCFKNVETASSEASDGA 283
 229 TISDNTREIYKKNNGCNGGAIQTSYTAIAKNNSGSVAFENNNATLSG--SINSNGSGCA 286
 284 IKVTRLDVTVNGRGRIFESDNTKNGGAIYAPVTLVNGPTFYNNIANNKGAIYID 343
 287 I-YTNILSIDNPGTILFNNNVCIRHGAICTOFTLKNSHYVFTNN-QGNMGALMLL 344
 344 GTSNKSISADKHAIIIFENIYNTVNTANGSTSANPRRAIITYASSSGEILLAGSSON 403
 345 QDSCTCLLFAEOGNIAFQNEVFLTFG-----RYNAIHCTPNNS-NQLQANNGYT 393
 404 LIFVDPIEVNAGVS-VFENKADOTGVSFSGATVNSADFHORNIOTKTPAPLTLNNGF 462


```
Db 80 LSTNSSGAFV--ESMSTFALDNADLFFCNVCTHOGG--GGAINAGLISFKNNQNTL 135
Oy 128 IFENNFCRPFSSPNNAVKKIRREGAIIHAQNTLYNHHDVVGFMKNSYRGALISTA 187
Db 136 FYNNITIOFTGVALKTERNR--GCALYGSSTIELINNHS--LNFINTNSGDMGCAVSTI 191
Oy 188 NTFVSENOGCLFMDNICIQ--TNT---AGKGAIIYAGTSNSESNNCDLFIINNACC 241
Db 192 ONLVYKNTSGIYAFENNHTDHIPTPATILARGAIVGCGACEISHNTGPVFNNSYCG 251
Oy 242 AGCAIIFS--PISCLTGNKGNIVFYNNRCKNVTASSEASDGAIKVTRTLDVTGNRGRIE 300
Db 252 YGCAISTGQCIFRDNKKLFIINNLSALGMHT--SAQNGAVISAGEFGLNNKGPY 309
Oy 301 FSDNITKNVGAIIYAPVTLVDNGPTFYFINNIANKGAIYIDGT--SNKISADRHAIIE 359
Db 310 FENNASTIAGISCNNLNFQENGPIYFLNNSA--LYGAFHLFASFAANYIHTGSGDIIF 368
Oy 360 NENITVNTNAN-----GTSTSANPPRRMAITVASSSGELLILGASSONLIFYDP 409
Db 369 NNNTLSTTGMAGLRKLEIYIGTINN-----NPITLS-----LGAKKOTRIIFYDL 415
Oy 410 IE-----VSNAGVSVFNKEADOTGSYFSGATVNSADFHORNIQYKT-----PA 454
Db 416 FOMGLKRAKPPENSPHTVITNPDSFGAVFYKYNI--SSDLQAHMIAKSTHNOIKDS 474
Oy 455 PTLTNGFLCIEDHAQ-----LVNRFOTGGVYSLNGAVLSCYKKGAGNSASNAI 507
Db 475 PTLTNGFLTSGIENGAEFFENGPILT-----QESTSLALQODSLITYGK-----DASL 522
Oy 508 TLKHLGLNLSILK--SGAEIPLWEP---TNSNNYATDRTATFSLSDVKL-----SL 557
Db 523 TITHGILPGLINDGQTAPRIARVNPQDMQNTNSNQ--PVSTENAVTOKIPEFSGLYSL 581
Oy 558 IDYGNSPRESIDLTHALLSSQPMLSISEASDQNRSD---DMDFSLNPPHYGMQCLW 614
Db 582 VQENESYVSDCLSRKANQPLILETNDQALSNDMKNLTMTLSLYSPHYGOGLMTS 641
Oy 615 GNAKTQDPPASSATITDOKANR-----638
Db 642 NMWTT--TPTVSLTNSTETQTNANISQEOBKNTSETPDSSTTAKIPSTRASTGTTPLA 699
Oy 639 -----FHRLLTLTWLPAGVSPKHSRPLIANTLWGMNLATESLKSNAELTPSDHP 692
Db 700 TTDVYTRISLIVSWPIPIYIADPARGLDINNVLVSSGNTLYLRS---LLPDD--SWF 755
Oy 693 GTTGGGLMMVYQDPRNHPGFMHSSGY--SAGMIAGOTHTFSLKFSQTYTKLNERIYAK 750
Db 756 ALQGSKATLFTQOKRLDYHGYSASASKGYAIISSQASGACHKFLFSPSSDDMKKERTN 815
Oy 751 NNVSANVSCQEGEMLFSLQEGELTLVYG--LYSYGDHNCHEFYTGENTSGTRESOT 808
Db 816 NKISSRYVL--SALCFEOMEDRIALIGAANVYGTHTKTYNFI--GTKKFSGKGNHSTT 870
Oy 809 MCGAVFFDL--PMKPFEGSTHILTPALGALIVSSLSHTEFGAVPPRSFTKPTLINVLP 867
Db 871 LGGSLGCLERDSMPQS--IMLPFIQALISRTREPASIOGQDLARLFLSKQHTAVVSP 928
Oy 868 IGVGSGFMNATORPQAWYELAYOPVLYROEPIATOLASKIHWGSGSPSSRHANYSK 927
Db 929 IGIKGYSSSKMPTVSCEMEVAYOPLLYMKRPLNLTLYLNNKGSWETNTPLAKHSF--YG 987
Oy 928 ISOOTQPLSWLTHFQYHGYSSTFCNYLNGEIALRF 965
Db 988 RGSSSLKFLTKLFLANTQAOVATSTVSHINAGCALVF 1025
```

```
RESULT 10
G71460
Probable outer membrane protein G - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
```

C:Accession: G71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: G71460
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1013 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; MID:g3329342; PIDN:AA68469.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmpG

Query Match 12.9%; Score 655.5; DB 2; Length 1013;
Best local similarity 25.7%; Pred. No. 2,le-32;
Matches 261; Conservative 152; Mismatches 388; Indels 213; Gaps 43;

```
Oy 39 SLNNKISLTGDNHNLNLCYLDNLRYLALILQKTPNCGAAVITDYLSFFDQKEGIYFAK 98
Db 84 NLGSPFLVGRGHSILT--FENIR-----TSTNGALS-----DSANGSLF--- 121
Oy 99 NLTPESGALGYASPNSPVLEIRDTIGPIYIFERNTCRREFTSNP--NAV----- 147
Db 122 --TIEGFKELFSNCSNLSLAVL---PAATTNNGSQTPPTTSTPSNGTIYSKTDLLIN 174
Oy 148 -----KIR--EGGAIIHAQNTLYNHHDVVGFMKNSYRGALISTAFTVSENOG 198
Db 175 NEKFSYSLNVLGSDGADIAKSLTYGICKLFCQENTQAOAGCAOVVTSFSAMANEAP 234
Oy 199 FLEMDNICIQTNTAG--KGAIIYA-----GTSNPF-----SNCDLFIINACCA 243
Db 235 IAFIANV-----AGVRGGIAAVQDQGVSSSTEDPVSFSRNTAVFEDGNVAVRG 288
Oy 244 GAIFSPICSLTGNKGNIVFYNN--RCEKN-----VERASSASDGGAI 284
Db 289 GGIYS-----YGVNAFLNNGKTLFLNNVASPVYIAAEOPITNGASANTSDNYGCGGAI 340
Oy 285 KYTRRLDVTGN-----RGRIFESDNITKNYGAIIYAPVTLVDNGPTFYFINNIAN 334
Db 341 FCNKQAOAGSNNSGSVSFEDGEGVFFSSVVAAGKGAIIYAKKLSVANCGPQFLOINIAN 400
Oy 335 NKGAIIYIDGTSNKSISADRHAIIFNENT---VTNVTANGTSTSANPPRRNAITVAS 390
Db 401 D--GCAIYLGESGELSLADYGDIIIFDGNLKRATAENAADVNGTVASS---QAIMSG 454
Oy 391 SGEILLAGASSONLIFYDPIEVN-----AGVSYFKNKEADOTGSYFSGATVNSADF 443
Db 455 GKITTRARAGHOLIFLNDPIEMANGNNOAPQSSSEPLKINDGEGYTDIYF--ANGNSTLY 512
Oy 444 HORNLOTKTPAPRLTSLNGFLCIEDHAQNTVNRFTOTGGVYSLNGAVLSCYKKGAGNS-- 501
Db 513 QN-----VTIEQGRVLEKAKLSVNSISOTGSGLYMAGSLDLDVTTPQPOQRP 562
Oy 502 ASNASITLKHIGLNLSSILKSGAEIPLWVEPTN--NSNNYATDRTATFSLSDVKLS--- 556
Db 563 AANQDLITLSLHLSLSSLLANNA-----VTNPPTNPAPRODHPAIIIGST--TAGSVTISGP 617
Oy 557 LIDDYGNSPRESYD-----LTHALLSSQPMLSISEASDQNLSDDDMDFGCLWVPHY 606
Db 618 FFEDLDQTAIDRYDWLGSNOK IDVLKIQIGTOP--SANAPSDTLT-----GNEPKY 667
Oy 607 GMOGLTWGAKTQDDEPASSATITDOKANRFRHTLLTWLPAGVSPKHSRPLIANT 666
Db 668 GYOGSKMLAN-----DPTNANNGPYTLAKATVTKTGTGNPGERARASLVNPS 712
Oy 667 LMGNNMLATESLKSNAELTPSDHPFW--GITGGGLGMVYQDPRNHPGFMHSSGY--SACM 725
Db 713 LMGS--IIDIRSAHSAIQASVDGRSYCGLWVSGVSNFFHYHDBRALQGYRYISGGSYLG-- 770
Oy 726 IACQTHFSLKFSQTYTKLNERIYAKNNV--SSKNYSQCGMGLFSLQEGFLTLTVLQGLYS 783
```

```

Db      221  FSONLAKSGGALYSTDNCSITDN--FOVLEFDGNSAMEAAQ-----AGCGAICCTTDTKT 273
Oy      290  LDVTGNRGRIFFSDNITKNGAIIYAPVVTLYVNGPFTYFINNIANK-----GGAIIYIDG 344
Db      274  VLTGNGK-NLSTNNMTALTYGCAISGLAKSISAGGPTLFQSNISGSSAOGCGGCAINIAS 332
Oy      345  TSNSKISADRAHAIIFENIIVTNANGSTSTSNAPPRNAITVASSGGEILGAGSSQNU 404
Db      333  AGELALSAISGDIENNNQVTN-----GSTST-----RNAINIDAKVTSIAMAQOSI 382
Oy      405  IYVDPIEVSNAQVS-----FNKEADQTSVVSAGATVNSAD-FHQRNLOTKRP 453
Db      383  YFVDP--TNPQTAOSTDTLNLNLADANSEIEGAIIVSEGEKLSPTFEALAAVSTIR 440
Oy      454  APLTSLNGFLCIEDHAQLTVNRFOTGGVVSGLNGAVLSCYKNGAGNSASNAITLKHIG 513
Db      441  OPAYILARGDLVLRDGVTVTFKDLTOSPGSRILMDG-----GTLISAKKAEALSLNGLA 492
Oy      514  LNLSSILKSGADITPLLWEPTNNSNNTADTAATSLSDVKKLSLIDDYNSPYESTDLTH 573
Db      493  VMLSSD-----DGTNKAAKTEADKNISLSTG-TALIDTEG-SEVENHNLSK 538
Oy      574  ALSSOPMLISASDN-QLRSDMDPSGLNV--HYGNGGLMTWGMKAKQDPEPASSATL 630
Db      539  A-STYPLELTTAGANGTTTICALSTLTLQEPETHYGYGNNQMSVAN-----ATSKK 591
Oy      631  TPOKANRHRFTLLTWLPAGYVPSPKHRSPLIANTLMGNMLLATESLKNKSAELTFSDHP 690
Db      592  GS-----INMTGTGIIIPPERKSNLPLNSLMGN-FIDISINOLETFKSSGP 638
Oy      691  F-----WGITGGGLCMVYVODPRENHDFHMRSSGYSGAMIAQOTHTFSLKFSQTYTKLNE 746
Db      639  FERELW--LSGIANFFYRDSMPTRHGRFHISGVALGITATPAEDQLTFA--FCQLRA 693
Oy      747  RAKANNVSKN-----YSCQGMELFSLOEGL-----LTKVLG-----780
Db      694  R-DRNHITTKNGDITYGASLYFHHTGELFDI-ANFLMGKATRAPVWLSLSQIIPISFDA 751
Oy      781  ---YSYGDHNCHEFTYOGENLTSOGTFRSOTMGAVFEDLPKMPFGSTHIL--TAPFLGA 835
Db      752  KRSYLTDHNMKTYT--DNSIIKSGMRNDAFCADGASLRF-VISVPLLKKEVEFPVKV 808
Oy      836  LCIYSSLSHTEVGAVPRSFSTKTPILNLVPIGVKGSFNNATQRPQAMVELAYQVVL- 894
Db      809  QYIYAHQODFYERYAEGRAFKN-KSELINVEIPIGVT---FERDSKSEKGTIDULMIYID 864
Oy      895  -YRQEGITOLLASKGIVFGSGSPSSRHMSKYSIQOTQPLSWLTLHFQYHGFYSSTF 953
Db      865  AVRNRKCTOTSLASDANMAYGTINLARGCFVYRANHFOVNPHEMLFGQF-AFEVRS 923
Oy      954  CNVLNGEIALRF 965
Db      924  RNV-NTNLSKF 934

```

RESULT 13

H81722
 polynucleotide membrane protein G family TC0263 [imported] - Chlamydia muridarum (strain N1)
 C.Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C.Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
 C.Accession: H81722
 R.Read: T.D.: Bunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBo, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A.Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A.Reference number: AB1500; M010:20150255
 A.Accession: H81722
 A.Status: Preliminary
 A.Molecule type: DNA
 A.Residues: 1-987 <TET>
 A.Cross-references: GB:AE002293; GB:AE002160; NID:g7190298; PIDN:AAF39132.1; PID:g719030
 A.Experimental source: strain N199 (Mopn)

C.Genetics:
 A.Gene: TC0263

Query Match 12.2% Score 620.5; DB 2: Length 987;
 Best Local Similarity 25.8% Pred No. 2.8e-30;
 Matches 253; Conservative 153; Mismatches 404; Indels 171; Gaps 42;

```

Oy      39  SLNKSISTGDTNHLNLCYLDNRLYLAILOKTPNEGA-----VITDLSFF 87
Db      82  NLGNFTINGRGHSLV---FENIR-----TSTGALSNHAPGLVIEAFDELSSL 130
Oy      88  DTQKECTFAKULTPESGAGIYASNSPTV-----EIRDITGPVIFENNTCCREPTSS 141
Db      131  NCNS-----LVSVPQGTGTT-TSVPSNGIYSRTDLVLRD-IKKVFSYNSVLS----- 177
Oy      142  NPMAAVNRIREGGAIHAQNDYINHNHDVYGFPMKRFYGVAGIASTANTFVSENOCELF 201
Db      178  -----GDCGALDAQSLMVGIEKLCTFDENVAOSDGCACQVTKTSANGNKVPLSF 228
Oy      202  MDNICIQTNTAGKGAIVA-----GTSNSFE-----SNCCDLFFINNACAGATFSPIC 251
Db      229  LGNV-----AGKKGGVAAVKDQAGAGATDLSVNFANNTAVFEENGSARIGGIIISD-- 281
Oy      252  SLTGNRGNIVFTNNR---CFKNVET-----SSEASDCAI-KYTRLD 291
Db      282  -----GNISFLGNAKTVFLSNVASPIYVDPAAAGQPPADKNDYDGGAIKNDTNIG 335
Oy      292  YVG--NRGRIFSDNITKNYGAIIYAPVVTLYVNGPFTYFINNIANKGAIYIDGTSNK 349
Db      336  EVSFKDEGVFFSKNIAACKGAIYAKKLLISDCGVQFELGNAAND-GAIYLVDOCELS 394
Oy      350  ISADRAHAIIFENIIVTNANGSTSNAPPRNAITVASSGGEILGAGSSONLFIYDP 409
Db      395  LSAHDRDIIIFDGLKMAITOGAATVHDV--MVASNAISMTAGQITTLRKRKEGRILFNP 453
Oy      410  IEVSNAG--VSSEFNKKEADQTSVFS-GATVNSADFHORNLOTPAPLTLNGLFCL 465
Db      454  IEMANGQPIQTLTVNEGEGYTGDIYFAKGDV-----LYSIEISQGRILL 500
Oy      466  EDHAQLTVNRFOTGGVVSGLNGAVLSCYKNGAGNSASNAITLKHIGLNLKLSGA 524
Db      501  REQTKLLVNSLTOTGGSVHMGSTLDF---AVTTPRANSMALTVNHFSLNLNNNV 557
Oy      525  ELPILVVEPTN-----NSNNYADTA--TFSLSDKVLSLIDYGNSPYESTDLTHALSSQ 578
Db      558  TNP-----PTNPVQVSPRAVIGNTAGTVTISGP--IFFEDETRAYDNNGMLGADQIT 610
Oy      579  PMLISASDONLRSDDMDPSGLNVPHYGMOGLMTWGMKAKTODPEPASSATITDPOKANR 638
Db      611  DVLQHLGANPRANAETDITLGNSSKXYGQGSMTLQW---EPDPA-----NPPONNS 660
Oy      639  FHRTLLTLTWLPAGYVPSPKHRSPLIANTLMGNMLLATESLKNKSAELTFSDHPV-GITGG 697
Db      661  Y--MLASMTKTGYNGPERVASLVNSLWGS--ILDVRSNHSIAQASIDGRAICRGIWIS 717
Oy      698  GLGMVYVODPRENHDFHMRSSGYSGAGMIA--GQHTFSLKFSQTYTKLNERAKNNV-- 753
Db      718  GINSEFFYHODALDAGQYRHISGYSIGANSYFGSS-MFGALFETETGR-----SKYVVC 771
Oy      754  SSKNYSQCGEMELFSLOEGFLTKLVG-----LYSGDNHCHEFTYOGENLTSOGTFRSQ 807
Db      772  RSNDDHCYVSUYSTQALGSCLEFDADAFVASYGNGOHMKTSYFAEB--SNVMDNN 829
Oy      808  TMGAVFEDLPKMPFGSTHIL--APFLAIGYSSLSHTEVGAVPRSFSTKTPILNLV 865
Db      830  CVYGEVAGAPRLAASKLYLNLPRVQAEFAVIAHESHTTEGDQAREKS-GHLMNLS 888
Oy      866  VPIGVKSEFNATQRPQAMVELAYQVPLYROEGIATOLLASKGIWFGSGSPSSRHMS 925
Db      889  IPIGVAFD-RCSSKHPNKKYSFMGAIYICDAVRSISGETTLLSHKEIWTDAFHLAHGWA 947
Oy      926  YKISQOTQPLSWLTLHFQYHG 946

```



```

OY 627 SATIIDPOKARHRTLLTLMPLAG---YVSPKHSRLPLANTLM--GNMMLA-TESEK 679
Db 663 DAKGMVPRNTN---NTLLYLMRPASNYGEYRLDOPKGELEPNLSLWAGSALRTFTNGIK 719
OY 680 -----NSAELTPSDHPFNGITGCGLMMVYODPRENNHPCGFHMRSRGYSAGMIAGOTH--- 731
Db 720 EHYVSRODGFVASLH-----ALGDYILANTQODDRGFLARYGFGQA---TASHYEN 768
OY 732 --TESLAKSSQTY--TKLMERYTAKNNVSSKNYSCOGEMFLSLDCEFLTKLYGLY--STGD 785
Db 769 GSIFCVAGGOLYGCQKRSRMYYSKDGANNMTMLSCFERSYVDIAG---TETVWYMETAG- 823
OY 786 HNCHEFTYOGENLTSGQTFRRSO-----TMGDA-----VFEDLPKMFEGSTHILFA-- 830
Db 824 YSVNHHMTQYFEDKTKQKRDHSHCKHHNNNNYAFUGAENHFLEKCIPTROFANDYDLTGEM 883
OY 831 PFLGALGIYSSHSFTEVGAYPRSPFTKPLINVLVPIGVKSGSPFNAYTOR--POAMTVEL 888
Db 884 RFEMAGGSSSSY---RELGSLTRYFARSGG--HNMSLRPLGIYVNAHVSIVHRSPRSKLTINM 939
OY 889 AYQPVLYKQEBGIATOLLASKGIWFGSGSPSRHMASTKXISQOTQPLSLMTLHFOYHGFY 948
Db 940 GYRPDIWRTVTPCNMELIANGVKTPIOGSPRLAHNAFFLEVHD-----TLYIHNFHGF-- 989
OY 949 SSSTFCNY 956
Db 990 --RAYMNY 995

```

RESULT 19
G81591
polymorphic membrane protein G family CP0303 [imported] - Chlamydomophila pneumoniae (Strain AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81591
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gall, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.: Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255
A:Accession: G81591
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-928 <REA>
A:Cross-references: AF:A002192; GB:AE002161; NID:g7189226; PIDN:AAF38160.1; PID:g718922;
A:Experimental source: strain AR39, HL cells
C:Genetics:
C:Gene: CP0303

Query Match	11.5%;	Score 586.5;	DB 2;	Length 928;
Best Local Similarity	25.6%;	Pred. NO. 3.2e-28;		
Matches 264;	Conservative 143;	Mismatches 377;	Indels 249;	Gaps 48

```

0Y 1 MKRAFFEEFLIGSL-----SGLREV-PSRIF-----LMPVSVDPTKRESIS 41
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKSOSWMLVLSLTACFTSCSTVFAPATAENICPDSOSFOSDSTNGTGTGTYPRNTTGTIDYTLT 60
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 42 NKISL--TGDTHNL--NCYLDNLR-----YIIAIIQ-KTPNCEAAVYIT----- 81
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 GDTTTONLSDSALITKGCSDPTTESLFEAGKGYSLFLINKSSAGGAALSVTTDKNLSLT 120
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 82 --DYSFFPQTQEGYFANLTPBESGAGYASPMSPYETIEDTIGPVFEENNTCCRPET 139
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 GFSLSLFLAPBSVY----TTPSGKAYKCGGDLT----FDNNGTILFKODYC----- 165
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 140 SSNPAAVANKIREGGAIHQNLNYINHHHDVYGFMKNFYSYVRGAISTANTFVVSQSCF 199
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 -----ERNGGAISTRKNLSLKNSTGISFEQKMS----- 193
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 200 LPMDNICIoTNTNAGKGAGIYAGTSNSESFNCCDLFFINN-ACCAGAGATSP--ICSLTGNR 257
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 -----SAGKKGAGCAATGATVDITNNTAPTLFSSNIIAAGAGAIINSTGNTTGTGT 244
    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

QY	258	GNIVFYNNRCKNVEVTASSESDGAIKVTRRLDVTNRGRGIFPSDNTKKYGGAIAPV	317
Db	245	-SLVFSSEN-----SVATATANGGALSGDDAVITISGMOQ-VTISGNOAVANGCAIYAKK	295
QY	318	VTLYVD-NGPTYFINNIAN----NKGAIVYIDGTSNKSISADRAHIIJENENIVTNTAN	371
Db	296	LTIASGGGGGJISFNNIIVQGTAGNGCAISILAGECSLSAEGADIFENGNAI-----	348
QY	372	GTSFSAANPRRNALITVASSGESEILLAGASSONLIFYDPIEVSNAGV---VSFK-----	423
Db	349	-VATTPQTTKRNSIDIGSTAKITNLRALISGHISFEYDEITANTADSDITLMLKADAGN	407
QY	424	EADQTSYVVFSGATVNS-----ADFHQRNLOTKTPAPRLTISNGEFCIEDHQLTVNRFQ	478
Db	408	STDPYSGSLVFESEKISDEBAKAVD---NLSTLKOPFTTLTACNLVILKRGVTLDTKGFQ	463
QY	479	T-GGVVSLGNGAVLISCYKNGAGNSASNAITLKHIGLNSILKSGAEPILLWEPTNNS	537
Db	464	TAGSSVIMDAGTTL-----KASTEETVLTLGISIPVDS-LDEGKKVVI-----	504
QY	538	NNYTADTATSLSSLYKIS-----LIDDYNSPREYETDL-THALSOPMLSISASDNL	591
Db	505	-----AASASAKKNVALSGPILLLDNOGNA-YENHDLGKTODESFVOLSALGTATTDV	556
QY	592	RSDMDFFSGLVNP-----HYGMOGLMTWGAKTODPEASSATITDPOKANFRHRTLL	645
Db	557	PA-----VPIYATPTTHYGIOSTMGMTVWDITASTPKT-----KTATL	593
QY	646	TWLPAGVYPSPKRHSPLIANTLMGNN-LTATES-LKNSAELTSPDHFMGITGGGLMM	702
Db	594	AMTNGTGLYLPNDRQGLPVNLSMGFSFDIQALQGVIERSAITLLCSDRFWM---AAGVANE	650
QY	703	VYQDPREHNPGEFHMRSSCYSAQMIAQOT---HTFSLAKSQTYYTLKNEBYAKNNVSKNYS	759
Db	651	LDKDKKGEKRKRRIHRSGGVYAIIG-AAQTCSENLISFAFCOLFQS---DDKDFLVAKNHT	704
QY	760	CGGEMLFSLQE-----GFLTLKLVG-----LYSGDGNCHNFT-GEEN	797
Db	705	DTYAGAFITQHTBESGCIICLLDKLPGSMWSHKPVLDEQOLATSHVSNIDLTKTITANREV	764
QY	798	LTSQGTFRSQTMGAVFDDLP--MKPFGSTHILTAPELFGALGIVSSLSHFTVEVAPRPF	855
Db	765	KGSMONNNAFMNMLIGASHSHYREYLHCPT---YAPYIKMLDUTYIRODSFEKGEGRSE	820
QY	856	SKTPTLIVLVLPICYK-CSFNANATORPOAMVVELAYOVULYROEFGIATOLLASKGIVFG	914
Db	821	DDSN-LFNLSLPIGYKVEKFEKSPDCND-FSYDLTISYVDDLIRNRPKCTTALVIGASMET	877
QY	915	SGSPSSRHAMSYK	927
Db	878	YANNILARQALOV	890

RESULT 20
D72067
polymorphic membrane protein A family CP0213 [imported] - Chlamydia pneumoniae (SP)
C:Species: Chlamydia pneumoniae pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text-change 11-May-2000
C:Accession: D72067; D81601
R:Kaitman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72067
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-947 <ARN>
A:Cross-references: GB:AE001638; GB:AE001363; NID:g4376819; PIDN:AD18679.1; PID:g4377
A:Experimental source: strain CML029
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000

```

OY 369 NANGSTSNAPRRNAITVASSSGEIL-IGAGSSONLITDPDIEVS-NAGVSFSFKEND 426
Db 351 SSSQTTT-----RNSINGNTNARIQOLRASGNTIYDPTTSTIALSDALNCP 404
OY 427 Q-----TGSVVFSGATVNSADPHOR-NLOKTPAPILNSGFLICIEDHADLTVRFTQ 478
Db 405 DLGAPNATOGITVIFSGEKISEAEADNLKSTIOQPLTLAGQSLKSGVTLVAKSFQ 464
OY 479 TCGVSLGNGAVLSCYKNGAGNSASNA-SITLKHIGLNLSSILKSGAEIPLLWEPFNN 537
Db 465 SPGSTLLMD-----AGTTLETADGITTINILVNLVNDL-----KETKKA 502
OY 538 NNYTADTATESLSDVKLSLIDDYG-----NSPYESTDLTHALSQPMLSIASON 589
Db 503 TLKATQASOTVTLTG-SLSLVDPSGCVYEDVSMNPNQVSCLTLLTADDPANHITDLAAD 561
OY 590 OLRSDDMDPSGLNVPHYGOGLTWGMMAKTQDPEPASATITDPOKANRPHRTLLTWLP 649
Db 562 PLEK-----NPIHWGYOGMMAISW---QEDTAKSKNAAT-----LTWK 597
OY 650 AGVPSPKHRSPLIANTLWGNML-----LATESLKSNAELTPSDHPFGITGGGLGM 702
Db 598 TGYNPNPERGTLVANTLNGSEFVDVRSIOQLVATKVRGQETR-----GIMCEGISN 650
OY 703 VYQDRENHPGFMHSSGYSAGMIAGQTH-----FSLKFSQTYTKLMERRA-KNNVS 755
Db 651 FHKDSTKIKKGRHISAGY-----VVGATTTLASDNLITAFCOLFKGDHFKNKRAA 706
OY 756 KNYSGOGEMLFSLQEGFLTKLVG-----LVSYGDNCHHFTYO---GEN-- 797
Db 707 YAASTLHLQHLATLSSSLRLYLPDSESEOPVLFDAQISITIKNTKTYTQAPKRESS 766
OY 798 -----LTSQGTFSQRTMGAVFIDLPMKPGSTHILTAAPLIGALGIS 840
Db 767 YNDGCALELASSLPHTALSHGELFHA-----YF-----PFIKVASYI 804
OY 841 SLSHFEVGA-ATPRSTKITPLINLVPIGVKGSFMNATQPOAMVELAYOVLVROEP 899
Db 805 HODSFERMTTLVRSFDS-GDLINSVPIGITYFERFSRMR-ASYATVYVADVYRKNP 862
OY 900 GIATOLLASGIVFGSGSPSRHA-----MSYKISQOTQPLSMLTLHFQ 943
Db 863 DCTTALLINTSKTGTGTLNSRQAGIRAGICITAFSPNLEVTNLSMEIR 912

RESULT 22
C72076
polymorphic outer membrane protein h family - Chlamydia pneumoniae (strain CML029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: G72076
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000. MIMD:99206606
A:Accession: G72076
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-978 <ARN>
A:Cross-references: GB:AE001629; GB:AE001363; NID:94376734; PIDN:AMD16596.1; PID:9437673
A:Experimental source: strain CML029
C:Genetics:
A:Gene: pmp_14

```

```

Query Match 11.5%: Score 583; DB 2; Length 978;
Best Local Similarity 26.1%: Pred. No. 5.6e-28;
Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
OY 41 SNNISLTGDRHNL--TNCYL-DNLRYLIALLOKTPNEGAATITIDVYSFDTQKEGIYFA 97
Db 67 SSNLSLGGKLSLFTTSCQAPTNSTVALL-----SAAETLTFKNF-----SINFT 112

```

```

OY 98 KNLPESSGAIGYASPNPTVEIRDTIGVIFENN-----TCCRPTSSNPNNAVKIREG 153
Db 113 GNSGTGLGLL-YGKD-----IVFOSIKDLITFTNRVAVSPASVTSATP-AITTVTTG 164
OY 154 G-----AIAHONLYINHHVDVGFMRKNEFSYRGCAISTANTFVVS--ENOSCELFMDN 204
Db 165 ASALQPTDLTYEN--ISGSIKFEGFLANF-----GSAISSPTAVAKFIMNATMSFSHN 218
OY 205 ICITQNTACKGCAIYAGTNSFESNNCDLFFINMCC-----AGALFS 248
Db 219 F-----TSSGGGVIVGSSLSLEFENNCGIIFPANSCVNSLKVTPSSGTYALGSGAICI 273
OY 249 P--ICSLTGNRGNIYFNNRCFKNVETASSEASDGAIVTTRLDVTGRGRIFESDNT 306
Db 274 PTGTLELKNQCKTFYSN-----GTPADCAIATET-CNIVGQCALLDSTNA 322
OY 307 KNYGCAIVAPVTVLVDNCPYFIJNIAKNGCAIYI-----DQTSNKSISADRHAI 358
Db 323 ARNGGALCAKVLNIGRGPIEFSRNRA-EKGAIFIGPSVDPAPAKOTSTLLASGDLA 381
OY 359 PNEIVTNTNANGTSTSNAPRRNAITVASSSGEILLAGSSONLITDPT-----E 411
Db 382 FQGNML-----NTRPGIRNAITVEAGGEIVLSAOGGSRVLVFDPTIHSLEPTTS 430
OY 412 VSNAGVSFENKADQTSVFSGATVNSAD-FHONLOTKTPAPILNSGFLICIEDHAO 470
Db 431 PSKKDITIAN--GASGAVFYSKGLSTELLIPANTTILLGTIKYIKSGELKITDNAV 487
OY 471 LTVNRF--TOTGVSLGNGAVLSCYKNGAGNSASNA-SITLKHIGLNLSSILKSGAEIPL 529
Db 488 VNLGFAITQSGGLTLGSGTLGLATPTGAPN--VDFTIGKLAFFPFSKLD----- 539
OY 530 WNEPTNNSNNTYADTATATSLSDVKLSLIDYGNSEPYESTDLTHALSQPMLSIASON 589
Db 540 FVASVYNA--TKNVTILTGAL-----VDLE-----HDVTDLYDNVSLSPVAIPTA-- 584
OY 590 QLRSDMDPSGLNV-----PHYGMOGLTWGMK--TQPEPASATITDPOK 635
Db 585 -----FRGATVTKTGPQDELATPSSHIGYOGKMSYMSRPLLIPADGCGPGSPS 636
OY 636 ANRHRITLITW-----LAGVPSPKHRSPLIANTLWGNMLATESLKSNA-----EL 684
Db 637 AN-----TLVAVMSDILVRSYTLDEPGEIYSNLSM-----ISFLGQARSDILOV 686
OY 685 TPSDHPFWGITGGGICGMVYQDRENHPGFMHSSGYSAGMIAGQTH--TFSLKESQTYT 742
Db 687 ILIDHGLSLITAKALGAYEHTRPGCHEGFGRYGGOALSNMYTDHTLLGSLFQOLY 746
OY 743 KLMERAKKNVSSKNY--SCOGEMLSLQEGFLTKLVGLYSYGDHCHHFTYQENL-T 799
Db 747 KTNANPYDSCHSOMVLSLFFGQPIVYTKSEALISKKAATYISKHNLNTYLRPDAPK 806
OY 800 SQCTFSQRTMGAVFIDLPMKPGSTHILTAPE-----FLGALGIYSSLSHFEVGA 850
Db 807 SOGOMHNNSYVLIAS--HFLNMCLLRLPLAQAMDLSGIFSAERLEGWQSKFTETGD 863
OY 851 YPRSFTKPLINLVPIGVKGSFMNATQ-POAMVELAYOVLVROEGIATOLLASK 909
Db 864 LORSFS-RGKGYNVSLPFGSSQWTFPKKAPSTLITIKLAKYKPDIVRVNHNITVVSNO 922
OY 910 GIVFGSGSPSRHAMSYSKISQOTQPLSMLTLHFQYHGFYSSTFCNY 956
Db 923 ESTSISGANLRHGLFVQI-HDVVDLTEDTQATLNTFTDCKNGFTNH 968

```

```

RESULT 23
B81593
polymorphic membrane protein h family CP0298 (imported) - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81593
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; McClarty, G.; Salabe

```

QY 455 ----PLTSLNGFLCIEDHQAOLTVNRFTQ-TGGVYSLGNGAVLSCKKNGASNASITLK 510
 DB 491 LNKLEGGTGLARNCATLVNHNFTQDEKSVYIMDAGTTLAT-TNCANN--TDGATLTLN 547
 QY 511 HIGLNTSLKSGAEIPLWVEPTNNSNNYADTAATFSLSDVKLSLDDYGNSESTED 570
 DB 548 KLVINLDSL--DGTKAAYVAVOSTNGA-----LTIISGLVLYKNSODCCDNHGM----- 594
 QY 571 LTHALSSQPMLSISEASDNLRSQ-DMDFSGLVNPHYGMOGLTWGAKTQDEPPASAT 629
 DB 595 FNKDQOVPILKATSNVTYTTDFSLGTNGCYQSPYGOGTWEF---TID---TTTHT 647
 QY 630 ITDPQKRNFRHRLTLTWLAGVPSPKHRSPLIANTLMGML-LATESLKSNAELPSD 688
 DB 648 VTG-----NMKKTGYLPHERLAPLIPNSLMANVIDLRAVSOASADGEDVP 694
 QY 689 HPFNGITGGGLGMVYODPRENHPGFHMRSSGYSGAGMIAGOTHT-----FSLKFSQTY 741
 DB 695 GKOLSTIG-----ITNFFHANNHG-DARSTRHMGGLINTYTRITPDALSLGFGOLF 747
 QY 742 TKLNERVAKNNVSKNYSQOGEMLFSLQEGFLTKLVGLSYGDHNGHFFYTOGENLT-- 799
 DB 748 TK-----SKDY-----LVG---HG-HSNVYFATVYSNITKS 774
 QY 800 ----SOGTFR-----SOTMGAVFFDLPKMPFGSTH 826
 DB 775 LFGSSRFPSGTSRVYTSRSNEKVKYTYKLPKGRCSMNSNCWLGELGNLPITL--SSR 832
 QY 827 ILT-----APFLGALGIYSSLSHFTVEGAYPRSFTKTPPLIIVLPDIGKGSFMAATORPQ 882
 DB 833 ILNLKQIIPFYKAEYAVATHGQIOENTPEGRIFG-HGHLVAVAVPVGRFG-KKSHNRPD 890
 QY 883 AMVELAYQPVLYROEPGIAIQTLLASKGIWFGSPSSRHAAMYKTSOOT 932
 DB 891 FYTIIVAVAPDVYRHNPDCTTLPLNGATWTISIGNLTLRSTLLVQASSHT 940
 RESULT 25
 C81593
 polymorphic membrane protein G family CP0299 (imported) - Chlamydomonas pneumoniae (stra
 C:Species: Chlamydomonas pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81593
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Morn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255
 A:Accession: C81593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-995 <RFA>
 A:Cross-references: GB:AE002191; GB:AF002161; NID:97189216; PIDN:AF8156.1; PID:9718922
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0299

Query Match 11.3% Score 575; DB 2; Length 995;
 Best Local Similarity 25.5%; Pred. No. 1.8e-27;
 Matches 273; Conservative 139; Mismatches 390; Indels 268; Gaps 50;

QY 1 MKKAAFFFLIGNSLGLAREVPSRIFLMPNSVPPTKESLSNKI-----SLTGDTHNLT-- 54
 DB 23 MKTSIRKFLIST--LAPCFASTAFTEVIEIMPSENFQSSGSKIPPTTLSPRGTLGIF 79
 QY 55 --NCYLDNLKRIILALLOKT--PNEGAVTITDYLSEFTQKEGIFAKNL-TPESGGAIG 109
 DB 80 SGDLVIANLDNAISRTSSCSFNAGALQILG-----KGVFSFLINRSSADGAAS 131
 QY 110 YASPSPIVEIRDTIG--PVIFENNTCCRPFTSSNPNAAVNRKREGAIAHONLYINHH 167
 DB 132 SVITONPELCLPSFSGFSQMTFDN---CESLTSD--TSASNVIPHASAIYATTPMLFTNN 186

QY 168 DVYGFMKNF-----YVKGAIANTFVYSENGSCFLFMDN----- 204
 DB 187 DSLIFQYNRSGAFGCAIRGISTILENT-----KSLFENGNSISNGALGSAIINLI 240
 QY 205 -----ICIQNTAG-KGCAIY-AGTSNFSFESNCDLFFINNACCAGAIFFPICSLTGNR 257
 DB 241 NSAPVIFSTNATGICGAIYLLTGSMILTSGLVLFVNNSSRSGAIA-----N 292
 QY 258 GNIVFYNR--CFKNVETASSEAS-----DGAIAKV-----T 288
 DB 293 GNVTFSSNLDLEQN-WTASFPQNSLPAPTPPPPAVTPPLIGYGAIFCTPPATPPPTGV 351
 QY 289 RLDVGNRGRIFESDNTIKNGAIYAPVYTLVNGCTFYFINNANKNGAIYDGTNS 348
 DB 352 SLTISGENS-VTFLENLASDEGALYCKKISIDSNKSTIFLGNTA-KGGALAIPESEL 409
 QY 349 KISADRAIIFENIIVNTNANGSTSANPRRNATIVASSGEILLGASSONLIYD 408
 DB 410 SLSANQDILEFNKL-----SITSGPTNNSIHFGKDAKFAILGATQGTLYFYD 459
 QY 409 PI-----EVSNAQSVSFNKED-----QTSVFSGATVNSADFHORNLOKTPA----- 454
 DB 460 PITSDDLASASAAATVYVNPKASADGAVSGTIVSEGLTAT-----EATPANATST 512
 QY 455 ----PLTSLNGFLCIEDHQAOLTVNRFTQ-TGGVYSLGNGAVLSCKKNGASNASITLK 510
 DB 511 LNKLEGGTGLARNCATLVNHNFTQDEKSVYIMDAGTTLAT-TNCANN--TDGATLTLN 547
 QY 511 HIGLNTSLKSGAEIPLWVEPTNNSNNYADTAATFSLSDVKLSLDDYGNSESTED 570
 DB 548 KLVINLDSL--DGTKAAYVAVOSTNGA-----LTIISGLVLYKNSODCCDNHGM----- 594
 QY 571 LTHALSSQPMLSISEASDNLRSQ-DMDFSGLVNPHYGMOGLTWGAKTQDEPPASAT 629
 DB 617 FNKDQOVPILKATSNVTYTTDFSLGTNGCYQSPYGOGTWEF---TID---TTTHT 647
 QY 630 ITDPQKRNFRHRLTLTWLAGVPSPKHRSPLIANTLMGML-LATESLKSNAELPSD 688
 DB 670 VTG-----NMKKTGYLPHERLAPLIPNSLMANVIDLRAVSOASADGEDVP 694
 QY 689 HPFNGITGGGLGMVYODPRENHPGFHMRSSGYSGAGMIAGOTHT-----FSLKFSQTY 741
 DB 717 GKOLSTIG-----ITNFFHANNHG-DARSTRHMGGLINTYTRITPDALSLGFGOLF 747
 QY 742 TKLNERVAKNNVSKNYSQOGEMLFSLQEGFLTKLVGLSYGDHNGHFFYTOGENLT-- 799
 DB 770 TK-----SKDY-----LVG---HG-HSNVYFATVYSNITKS 774
 QY 800 ----SOGTFR-----SOTMGAVFFDLPKMPFGSTH 826
 DB 797 LFGSSRFPSGTSRVYTSRSNEKVKYTYKLPKGRCSMNSNCWLGELGNLPITL--SSR 832
 QY 827 ILT-----APFLGALGIYSSLSHFTVEGAYPRSFTKTPPLIIVLPDIGKGSFMAATORPQ 882
 DB 855 ILNLKQIIPFYKAEYAVATHGQIOENTPEGRIFG-HGHLVAVAVPVGRFG-KKSHNRPD 912
 QY 883 AMVELAYQPVLYROEPGIAIQTLLASKGIWFGSPSSRHAAMYKTSOOT 932
 DB 913 FYTIIVAVAPDVYRHNPDCTTLPLNGATWTISIGNLTLRSTLLVQASSHT 962

Search completed: May 6, 2001, 19:18:48
 Job time: 543 sec

546ThrPheSerTrpLysAspSerAspGluGly.....HisSerL 558
 668 TCAGTGTACAGGACAGAGGACAGATGATGATGCTGCTCCACCTCTC 717
 558 euIIeAlaAsnTrpThrProLysAsnTrpValProHisProGluArgL 574
 718 GCCTCTGGGACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 767
 575 SerThrLeuValAlaAsnThrLeuTrpAsnThrTrpSerAspMetGln 591
 768 CCCATCTGCGACGCTCA.....AAAGCGTTTGGC...AAAGC 805
 591 aValGlnSerMetIleAsnThrIleAlaHisGlyAlaTrpLeu.... 606
 806 CTGTCTACAGAGGTCAGAGTATGTCAGCGCTCTCCCTCAGTGCACC 855
 607PheGlyThrTrpGlySerAlaValSerAsnLeuPhe 618
 856 CAGTCAACCCCTTAGATGCACACGCTGGGCTCATGTTGTAATATATAT 905
 619 TyrAlaHisAspSerSerGlyLys.....ProIleAspAsnTrpHis 632
 906 TCACGATATTCCTGTAACACAGGTTTGTGCTCTAGCTAATTTCAT 953

seq_name: gb_gss30:CNS02ESG

seq_documentation_block: 877 bp DNA GSS 13-MAY-2000
 LOCUS CNS02ESG Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
 DEFINITION 131E19 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL194137 GI:7832243
 VERSION 1
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
 Holacanthopterygii; Acanthopterygii; Percomorphi;
 Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 877)
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
 Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis

TITLE Unpublished
 JOURNAL 2 (bases 1 to 877)
 REFERENCE Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence

TITLE Unpublished
 JOURNAL 3 (bases 1 to 877)
 REFERENCE Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES
 source
 1..877
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="131E19"
 /clone_1lb="G"
 /note="Genoscope sequence ID : C0AG11AC10SP1-end :
 PUC-ori"

BASE COUNT 131 a 298 c 127 g 318 t 3 others

ORIGIN

alignment_scores: 115.00 Length: 346
 Quality: 0.676 Gaps: 15
 Percent Similarity: 49.133 Percent Identity: 21.676

alignment_block:
 US-09-677-752-4 x CNS02ESG

Align seg 1/1 to: CNS02ESG from: 1 to: 877

206 lIecyScySerAsnLeuIleCyS..... 213
 22 ATTTGTTGGCGCATCTCTAATGCATGATACACTTCATCTTCACCT 71
 214SerGlyAsnValAsnProLeuPhePheThrGlyAsnSerAla 228
 72 ATCTTCTTCTCTCTGCTGCGCCAAAGAGATTTTCTTCTCATTCATCTT 121
 228 hrAsnGlyAlaAlIecyScyIleSerAspLeuAsnThrSerGluLys 244
 122 CTGCTGCATCTGCTCTTCATCAGTACAGAGAGTCTCTCTCCACCTCC 171
 245 GlycerLeuSerLeu.....AlaCy 251
 172 TCCTCTCTTCTGCTGTGCGCATCACCCTCATCTTCCTGATTCCTG 221
 251 sAsnGlnThrLeuPheAlaSerAsnSerAlaLysGlyLysGlyAla 268
 222 CTGGATGAAGCTGCTCAGCGGCTCTCTCTGTAACCTCC..... 264
 268 laIeTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSerPhe 284
 265TCCTCTCTCCGCGCTTC 282
 285 lIeAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerGlyAl 301
 283 CTCTCATCTTCTCT.....TCAGGTTCTCTGCTCTCTCTGCGAG 323
 301 ySerLeuSerIleLeuAlaGlyGlySerValLeuPheGlnAsnAsn 318
 324 AGCATCTTCATTTGTGAGATCGATTTTTCATCTTGTCCACCTTCT 373
 318 erGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuGlu 334
 374 CCTCTGCTGCCAATGAC.....CTGCAG 396
 335 LysAspAlaIleLeuSerSerLeuGlnAlaArgAsnGlyAspIleLeu 351
 397 CAACCTGCCGCGGCTCTCTCC..... 417
 351 ePheAspProIleValGlnGlnSerSerSerLysGlnSerProLeuPro 368
 418TCTCATCA.....T 427
 368 erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
 428 CAGGACTCTCTCTCTTTCGCCCATCATCTTCACCTGATACCTGCTCAG 477
 385 leuValIleGlnThrSerAlaAsnArg..... 393
 478 GAAGCTACTCACTAGCTCTCATCTTCGCTTGAAGAAGATCTTTT 527
 394SerValIlePheSerSerGluArgLeu..... 402
 528 TCTGACGTTTCATGTCAGTTGTGCTTCTTCATCAGCTGATGCTGAT 577
 403SerGluGlnLysThrPro 409
 578 CATCGTTTCCATCTGATTTGTGATTCAGATTCTTCTTCATCAGACAGAT 627

```

394 SerValIlePheSerSerGluValGluSerGluGluGlyThrProAs 410
      |||:
275 TTAGTATTATA.....CCTTC 259
410 PAsnLeuThrSer.....GlnLeuGlnInProIleG 421
      |||:
258 AAGCTTACTACTCTTCATTAATAATGTTCAATTGAAAAAGTCACTG 209
421 IuLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAla 437
      |||:
208 AAACCACTTCAGGCTCTTACTCCTTAACCA.....ACTGTT 171
438 ProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyTh 454
      |||:
170 ACAAGCTCTTCTATTGAAAA.....ATAGAAACTGTTTC 136
454 rSerLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIleProL 471
      |||:
135 TTCTTTAAACACA.....ATCCCTT 116
471 euHisSerLeuAspThrGluLysSerValThrIleHisAlaProAsnLeu 487
      |||:
115 TATTTATTTCAGAACT..... 99
488 SerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTyG 504
      |||:
98 .....TCAGAAATACCAATTCTATTAGTGTAT..... 69
504 uAsnValGluLeuLeuSerLysGluGlnAsnAsnIleProLeuLeuThrL 521
      |||:
68 .....CTTCATCAGTTTCAATAATAATTATTATTACTA 32
521 euSerLysGluGlnSerHisLeuHis 529
      |||:
31 AAAGTAGAAAACTCTCTTCCAGCAC 6
seq_name: gb_es193:BF864732
seq_documentation_block:
LOCUS BF864732 741 bp mRNA EST 19-JAN-2001
DEFINITION 963054D01.x1 C. reinhardtii CC-1690, stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF864732
VERSION BF864732.1 GI:12254876
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 741)
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stein, D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Cellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES
Source location/Qualifiers
1..741
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 219r"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, Lambda Zap II"
/note="vector: pluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; This library, constructed by John Davies and Jeffrey

```

```

McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
zap clones by superinfection with EXassist (Stratagene)
phase. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 289 a 32 c 383 g 37 t
ORIGIN
alignment_scores:
Quality: 109.50 Length: 211
Ratio: 0.898 Gaps: 11
Percent Similarity: 57.820 Percent Identity: 29.384
alignment_block:
US-09-677-752-4 x BF864732/rev ..
Align seg 1/1 to reverse of: BF864732 from: 1 to: 741
339 LeuSerSerLeuGluAlaArgAsnGlyAspIleLeuPhe.....PheAs 353
|||||
702 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 665
353 ProIleValGlnGlnSerSerSerLysGluSerProLeuProSerSerL 370
|||||
664 CTCGATACCTCTCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 615
370 euGlnAlaSerValThrSer...ProThrProAlaThrAlaSerProLeu 385
|||||
614 TCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 565
386 ValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLe 402
|||||
564 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 535
402 uSerGluGluGluLysThrProAspAsnLeuThrSerGlnLeuGlnInP 419
|||||
534 .....TCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 498
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeu 435
|||||
497 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 448
436 Ser.....AlaProSerLeuSerGlnAspProGlnAlaLeuLeuI 449
|||||
447 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 398
449 eMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla. 465
|||||
397 TCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 348
466 ThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValThrI 482
|||||
347 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 307
482 euHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyA 499
|||||
306 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 260
499 spGluAsnPheTyGluAsnValGluLeuLeuSerLysGluGlnAsnAsn 515
|||||
259 CTCTCCCTCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
516 IlePro...LeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
|||||
224 CTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175

```

COMMENT

Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES

source

Location/Qualifiers
1. 1166
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI, Site_2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
Zap clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 419 a 20 c 506 g 197 t 24 others
ORIGIN

alignment_scores:

Quality: 106.00 Length: 273
Ratio: 0.785 Gaps: 12
Percent Similarity: 49.451 Percent Identity: 24.542

alignment_block:

US-09-677-752-4 x BF864174/rev ..

Align seg 1/1 to reverse of: BF864174 from: 1 to: 1166

```

349 1leuPhepAspProIleuValGlnIuSerSerSerylSgIuSerPr 365
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
988 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 939
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
365 oleuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrA 382
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
938 TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 889
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
382 laSer.....ProLeuValIle 387
      ||| :::::::::::
888 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 839
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
388 GlnThrSerAlaAsnArgSerValIlePheSerGlnAluGluSerG1 404
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
838 CCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTACT 789
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
404 uGluGluysThrProAspAsnLeu.....ThrSerGlnL 416
      ::::::::::: :::::::::::
788 TTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCC 739
      ::::::::::: :::::::::::
416 euGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg 432
      ::::::::::: :::::::::::
738 TGCACTGCTGCACTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 689
      ::::::::::: :::::::::::
433 AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuL 449
      ::::::::::: :::::::::::
688 TCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCTT 639
      ::::::::::: :::::::::::
449 eMetGlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysAlaT 466
      :::::::::::
638 CTCC.....TNCCTCTTAC 625

```

seq_name: gb_est82:BF064737

seq_documentation_block:

LOCUS BF064737 1688 bp mRNA EST 17-OCT-2000
DEFINITION HY_CBD0017K08f Hordeum vulgare seedling green leaf EST library
HY_CBDNA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
HY_CBD0017K08f, mRNA sequence.
ACCESSION BF064737 GI:10841376
VERSION BF064737.1
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae;
Hordeum.
REFERENCE 1 (bases 1 to 1688)
AUTHORS Wing, R., Close, T.J., Kleinborts, A., Wise, R., Begum, D., Frisch, D., Yu
Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics
JOURNAL Unpublished (2000)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: ATTATACCTCTCAAGCG
High quality sequence start: 65
High quality sequence stop: 1132.
Location/Qualifiers
1. 1688
/organism="Hordeum vulgare"
/cultivar="C116155 (Ma13)"

FEATURES
source

```

466 hrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValThrIle 482
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
624 TCTTCTCACTACACACTACACACTACCTCTTCTCTCTCTCTCTCTCT 575
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
483 HisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAs 499
      ::::::::::: ::::::::::: ::::::::::: :::::::::::
574 TACTGCCCATCTCTCAGTC...TNCCTCTTC..... 545
      :::::::::::
499 pGluAspPheTyrlGluAsnValGluLeu..... 509
      :::::::::::
544 ...TCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 499
      :::::::::::
510 .....SerLysGluGlnAlaAsnIleProLeu..... 519
      :::::::::::
498 GCTACTATCATCTTATCTTATTAATCTCTCTCTCTCTCTCTCTCTCT 449
      :::::::::::
520 ThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGlyAsnLeuSe 536
      :::::::::::
448 CTACTACTCTCTCTCACTCTCTCTCTCT...CTCTCTCTCTCTCTCT 402
      :::::::::::
536 rSer.HisPheGlyTyrlGlnLysAspThrPheSerThrLysAspSer 552
      :::::::::::
401 CTCTCATCTCTACACTNCAACNCACTCTCTCTCTCA..... 364
      :::::::::::
553 AspGluGlyHisSerLeuIleAlaAsnThrPheProLysAsnTyrl.... 567
      :::::::::::
363 .....TCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTAC 326
      :::::::::::
568 .....ValProHisProGluArgGlnSerThrLeuValAlaAsnThrL 582
      :::::::::::
325 TCTACCACTCACTCAACAACTCACTCACTCACTCACTCACTCACTCA 282
      :::::::::::
582 euTPAsnThrTyrlSer 587
      :::::::::::
281 TATACTCACTCTCTCA 265

```


291 eglygylalailealieglnserglyglyserleuSerlleuAlaG 308
 462 ATCGGGAAGACTCTCG.....GGATTGAACTCTCGGCAAGCC 499
 308 lylgluglyserValleupheglnAsnAsnSerGlnArgThrSerAspGln 324
 500 GAGAAAGCAGCGCTTACCTTTTGAGGAGGAGTCTTTTACCGACTCTCC 549
 325 GlyleuValArgAsnAlaIleTyrLeuGlnLysAspAlaIle..... 338
 550 GGGGACAGCTCAGCGCTCGAGCTTGAGACAGGATCGGTGGAGTACAG 599
 339LeuSerSerLeuGlnAlaArgAsnGlyAspI 349
 600 CCGGAGAGGAAAGCGGACAGCACTGCTCTCCACTCGCCGAGCAAGACA 649
 349 lEleupheAspProIleValGlnGlnSerSerSerLysGlnSerPro 365
 650 ACCGATGTCGAGCAGCGGAGGAGCTCTCGGACTTCTCTCGCAGAC 699
 366 LeuProSerSerLeuGlnAlaSerValThrSer.....ProThrPr 379
 700 CTGCGGAGAGTGTCTGCTCCTCCCTCCACTCTCGAGTGGCAGCGACTGA 749
 379 oAlaThrAla 382
 750 GAGAAAGCGCC 759
 seq_name: gb_est90:BF623271

seq_documentation_block: 1541 bp mRNA EST 18-DEC-2000
 LOCUS BF623271 HVSMBA0016E13f Hordeum vulgare seedling shoot EST library
 DEFINITION HVCDNA001 (Cold stress) Hordeum vulgare cDNA clone HVSMBA0016E13f,
 mRNA sequence.
 ACCESSION BF623271
 VERSION BF623271.1 GI:11887101
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota: Viridiplantae; Embryophyta: Tracheophyta: Spermatophyta:
 Magnoliophyta: Liliopsida: Poales: Poaceae: Pooidae: Triticaceae:
 Hordeum.
 1 (bases 1 to 1541)
 Wing,R., Close,T.J., Kleinholz,A., Wise,R., Begum,D., Frisch,D., Yu
 ,T., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
 ,T., Saski,C., Schwartzbeck,J., Simmons,J., Chol,D.W., Main,D. and
 Wood,T.
 Development of a genetically and physically anchored EST resource
 for barley genomics
 Unpublished (2000)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: AATTAACTCTCACTAAAGG
 High quality sequence start: 79
 High quality sequence stop: 967.
 location/Qualifiers
 1..1541
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMBA0016E13f"
 /clone_lib="Hordeum vulgare seedling shoot EST library
 HVCDNA001 (Cold stress)"
 /tissue_type="Seedling shoot"
 /lab_host="TJC121"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 338 a 748 c 30 g 407 t 18 others
 ORIGIN
 Alignment_scores:
 Quality: 103.50 Length: 267
 Ratio: 0.877 Gaps: 12
 Percent Similarity: 44.195 Percent Identity: 24.345

Alignment_block:
 US-09-677-752-4 x BF623271 ..
 Align seg 1/1 to: BF623271 from: 1 to: 1541

351 phepheaSproIleValGlnGlnSerSerSerLysGlnSerProLeuPr 367
 513 TTCTCAAACTCTACTACACACTACACCCCATACACACTCCACTCC 562
 367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
 563 CTCACACACACACTCCACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 612
 384 roLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer..... 398
 613 CATCTCTCTC.....TCACCATTCATTCACACTTATTCACCTATNTC 656
 399SerGluArgLeuSerGlnI 405
 657 TCTTCCCTATATACACTCTTTTACATATCAACAACCATGGGGGAGAGA 706
 405 uGlnLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGlnL 422
 707 CCCCACACTCT.....CATCCCTCTTAACCTC 735
 422 euLysSer.GlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPr 438
 736 TCTTTCTCTCAGCTCCTCTCCACACACTCTCTCTTATCTCAACAC 785
 438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGlnAlaGlyThrS 455
 786 ANTA.....CCACCCCTCTTCACACACATACACACACT. 822
 455 erLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIleProLeu 471
 823ACCTCTCACCCCTTCCCTCCCTCTCTCTCTCTCTCTCTCTCT 864
 472HisSerLeuAs 475
 865 AACTCTCAAAAACAAATCAAAATGCTATACCCCTCTCTCTCC 914
 475 pThrGlnLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysI 492
 915 CTCACACT 964
 492 lEphLeuSerAsnSerGlyAspGlnAsnPheTyrGlnAsnValGlnLeu 508
 965 TTATCTCTCTCTTTATCATATACAC..... 990
 509 LeuSerLysGlnGlnAsnAsnIleProLeuLeuThr.....LeuSerLys 523
 991 CTATACACCAACTCTCTCATCTCTCTGACTACCTCTCTCTCTCTCTCT 1040
 523 sGlnGlnSerHisLeuHis.....LeuProAspGlyAsnLeuS 536
 1041 TCCCT 1090
 536 erSerHisPheGlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSer 552
 1091 CCACCAT.....CACTCTTCC 1107
 553 AspGlnGlyHisSerLeuIleAlaAsnTrpThrPro..... 564
 1108 GACTATTCACACACACACCTCTATCTACCTCCCTCTCTCTCTCTCTCAN 1157

and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGACGCGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 383 a 527 c 302 g 253 t 1 others

alignment_scores:
Quality: 101.00 Length: 344
Ratio: 0.623 Gaps: 18
Percent Similarity: 47.093 Percent Identity: 23.837

alignment_block:
US-09-677-752-4 x BF570499

Align seg 1/1 to: BF570499 from: 1 to: 1466

```

366 leuproSerSerleuGlnAlaSerValThrSerProThrProAlaThrAl 382
|||||
333 CTACCCAGTCCACACAG...TCTCTGACGACCCACGCTACTGGGCT 379
382 aserProleu.....ValIleGlnThrSerAlaAsnArgSerValI 396
|||||
380 GCTTCACCTTACGCCCTATCCTGTATACACAGCACCCTTGGAAGCCCG 429
396 lephSerSerGluArgSerSerGluGluGluValThrProAspAsnLeu 412
|||||
430 TTATC.....AACTCTGCTGATG 452
413 ThrSerGlnLeu.....GlnGlnProIleGluLeuLysSe 424
|||||
453 ACTTCCACGTGTACCCATCCTGTGACACGCCACCTGTGACTGACTG 502
424 r.....GlyArgLeuValLeuLysAspArgAlaValLeuSerA 437
503 CACCGCACATTACGACCACTCTGGTAAGCTGTCACGCCCTGCTGG 552
437 laphSerLeu.SerGlnAspProGlnAlaLeuLeuIleMetGluAlaG1 453
|||||
553 CACCACACTTGGAGCAAA...CGCTGGCCTTGTTCGATACAGTACGTCG 599
453 yThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLeuSerIle. 469
600 C.....CCAACTGGTACACCACTCAACCTGCTAGTACAGGTATCC 640
470 ..ProleuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro 485
641 TTCCACAAACACATCTACAGCCCAATACACCTACATGAGGGGCCACA 690
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspS1 500
691 ATGGGGCGGCTTGGCATTTGGCCCATTTACAAACGGGGCCCCAGACAGAC 740
500 uAsnPheTyGluAsnValGluLeuLeuSerLysGluGlnAsnAsnIleP 517
741 AAATTCACATCCAAACAGCCAACTTATTCGTGGGAGAACAGCGCATACC 790
517 roleuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGly 533
791 CTGCACAGGTCCACAGCGGCTGTGGCCACCATTTCCACACACCT..... 834
534 AsnLeuSerSerHisPheGlyTyGlnGly.AspTrp.....ThrPheS 548
835 CGACATTGGACGGCTCAATGGACAGCGCTTTCATGGACAGAAATTTCT 884
548 eTrpLysAspSerAspGluGlnHis.....SerLeuIleAla 560
885 CATGGCATACTAGGGTGGCCCAACACACACCCCTTTGGCGCTCGCAAGC 934

```

```

561 AsnTrpThrProLysAsnTyValProHisProGluArgGlnSerThrIle 577
935 AACACCCCCCGAGGGCTCGTGTCCCTACACCATCAACAAACAAAGCT 984
577 uValAlaAsnThrLeuTrpAsnThrTySerAspMetGlnAlaValGlnS 594
985 TTCACCAACA.....CATACAACTTTCATACCGTCATGGGGGT 1025
594 ermelleAsnThrIleAlaHisGlyAlaTyLeuPheGlyThrTrp 610
1026 GCTGATTAGACCACTCGCAGCATGGCGGACAGCCTCAATGGGACAA... 1072
611 GlySerAlaValSerAsnLeuPheTyAlaHisAspSerSerGlyLysPr 627
1072 ..... 1072
627 oLleAspAsnTrpHisHisArgSerLeuGlyTyLeuPheGlyIleSer 644
1073 .....AACTACCCCGGACAGCCTTA.....ACAA 1097
644 hrHisSerLeuAspAspHisSerPheCysLeuAlaIleGlyGlnLeu 660
1098 CAGATGGCCAGACAGACACCATCGTCCGTCGCGAGGGGACG..... 1141
661 GlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTyrl 677
1142 .....ACGAAACAATCCACCTATCG 1161
677 eAlaThrValGlnAlaGlnLeuAlaThr 686
1162 CACACACACCGGGGCAACCGGCCACA 1189

```

seq_name: gb_est49:AW694070

seq_documentation_block:

LOCUS AW694070 665 bp mRNA EST 15-JUN-2000

DEFINITION NF072B11STP1092 Developing stem Medicago truncatula cDNA clone

ACCESSION NF072B11ST 5', mRNA sequence.

VERSION AW694070 GI:7568807

KEYWORDS

SOURCE

ORGANISM

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

Medicago truncatula

229	CGTGGGCGGAGCAACGTCGAGGTGTC...ATCGATGTCCTGGGCGGC	345
163	PheThrAlaPheGluGlnAsnSerLysGlyAsnGlyValAlaIleI	179
346	TTCAACCCACATGAGAGAGGCCCATCTCCACACAGAGGGTGGTGAAGA	395
179	nalaglnThrPheSerLeuSerArgAsnValSerProIleSerPheAla	196
396	GGTCCTGATCACCAGCCCGCTGGGCGGACAGCCCATGTTGCTGACGGT	445
196	rgAsnArgAlaAspLeuAsnGlyValAlaIleCysCysSerAsnLeuIle	212
446	TCAACCCACAGTCGATCAAGCGGACAGAGGGTCTTCTGTAACGGCTCG	495
213	CysSerThrLysAsnValAsnProLeuPhePheThrClyAsnSerAlaThrAs	229
496	TGCACGACCAAC.....	507
229	nglGlyValAlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlyS	246
508TCCCGGGGGCCCTCCACCAAGCATCATCAGCACA	541
246	erLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAla	262
542	AGTTCCGATCGTCCGGCGCTGATGCACCCGTCACATCGACAGCGCC	591
263	LysGluLysGlyAlaAlaIleTyrAlaLysAsnMetValLeuArgTyrAs	279
592	ACCCAGAG.....ACCGTCA	608
279	ngLysPro.....ValSerPheIleAsnAsnSerAlaLysIleGlyAla	294
609	CGGCGCCCTCGGCAAGCGCTGGCCGACAGCGCGTGGCGCGCGCGCACA	658
294	IaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaGlyGlyGly	310
659	TCATCCCTCGACGATCGGCCCGCCACAGGCGTC.....	693
311	SerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeuVa	327
693	693
327	IaArgAsnAlaIleTyrLeuGlyLysAspAlaIleLeuSerSerLeuGlnA	344
694ACGCCGTCATCCCGACGTCACAG	718
344	IaArgAsnGlyAspIleLeuPhePheAspProIleValGlnGluSer..se	360
719	GGCGGCTACAGTGTGCTTCGGCGGCTCCCGACGCTCAAGCTGCTGCTC	768
360	rSer.....LysGluSerProLeuProSerSerLeuGlnAlaAs	373
769	GTGACACGTGACTGCCGAGCTCAAGCCCGCCACGTACAGAGGAGTCTG	818
373	erValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr	389
819	CAACCCGCTCGTGAGCGGCCCAAGCGGCGACTCA.....AGGGCG	859
390	SerAlaAsnArgSerValIlePheSerSerGlyLysArgLeuSerGluGluI	406
860	TCCTCGGCGTACACCGACGAGCGTGTCTGTGCGAATCTGTCGCGGAG	909
406	uLysThrProAspAsnLeuThrSerGlnLeuGlnInPro...Ile..Glu	421
910	CCCATGTCGTCGATCTTGACAGCCCAAGCGGGAATCGCCCTGCACAAAC	959
422	LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeu.....	435
960	CTTTCGTCAGACTGGTCTCTGTGACACACAGAGTGGGGCTACTTCGAC	1009
436SerAlaProSerLeuSerGlnAspPro...GlnAlaLeuL	448

Email: rwing@clmson.edu
Seq primer: ATTTAACCCTCACTAAGGG
High quality sequence stop: 1831.
Location/Qualifiers
Source
1. 1831
/organism="Hordeum vulgare"
/cultivar="Cili6151 (Mla6)"
/db_xref="taxon:4513"
/clone="HY_CBD006019f"
/clone_110="Hordeum vulgare seedling green leaf EST
library HVCDNA0005 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 573 a 184 c 678 g 351 t 45 others
ORIGIN
Alignment_scores:
Quality: 99.00 Length: 446
Ratio: 0.497 Gaps: 18
Percent Similarity: 44.619 Percent Identity: 21.300
Alignment_block:
US-09-677-752-4 x BE215233/rev ..
Align seg 1/1 to reverse of: BE215233 from: 1 to: 1831
354 ProilevalingluserSerSerlysgluserProleuProSerSerle 370
1804 CCGGCGCCGCCACCTCCANTCTCTCCACACGCCCTCACTCCATTAC 1755
370 uGlAlaSerValThrSerProThrProAlaThrAlaSerProleuValI 387
1754 TACTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATTTC 1705
387 LegInThrSerAlaAsnArg..... 393
1704 CTCACACCCACCCACATCACACACACGACGACCGCTTTGTAG 1655
394SerValIlePheSerSerGluArgLeuSerGluGluGlu 407
1654 AGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1605
407 sThProAspAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 424
1604 ATGTCACATATATCCACT.....CCACTCCGCTCTCCATCACACA 1564
424 erglArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLeu 440
1563 CTCACCGGCTGCGCTACCC.....CTCTCTCTCTCTCTCTCTCT 1526
441 SerGlnAspPro..... 444
1525 CCATCACTCCGCTGCTGCTGCTTTATACCTGCGCTTATCAGATT 1476
445GlnA 446
1475 GATTGTCTCTGCATATTGTTTATCTTGTACATTGATATTCTTCAATT 1426
446 lAlaLeuIleMetGluAlaGlyThrSerLeuLys..... 457
1425 TCTTATTGATCATATTATACATATTCTCTCTCTCTCTCTCTCTCTCT 1376
458 ThrSerSerAspLeuLysLeuAlaThrLeuSerIleProleuHisSerI 474
1375 ACCTCTCTCACTGCTTCATCGCGGCTCTCTCTCTCTCTCTCTCTCT 1326
474 euAspThrGluLysSerValThrIleHisAlaProAsnLeu..... 487
1325 TCCCTTCACTCTCACTCACTCACTCACTCACTCACTCACTCACTTAC 1279
488 SerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPhetyrGI 504

11278 TGTGACTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1249
504 uAsnValGluLeuLeuSerLysGluGlnAsnAsnIleProleuLeuThrL 521
1248CTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1203
521 euSerLysGluGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSe 537
1202 TCTCTCTCGCGCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCN 1153
537 rHisheGlyTyrGlnGlyAspThrPheSerThrPlyAspSerAspG 554
1152 GTCCTATCCGCTC.....NTG 1136
554 lUGlYHis...SerLeuIleAlaAsnTrpThrProLysAsnTrpValPro 569
1135 TGTCTCATCTCCAGCTGTACTAGCCCTGTCTAGCCATCAGCTGTCTCA 1086
570 HisProGluArgGlnSer..... 575
1085 CATATTCTCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1036
576ThrLeuValAlaAsnThrLeuTrpAsnT 585
1035 TCTTCATCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 986
585 hrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHis 601
985 CTTACTCTCTCTGACTTTATTAATATCATCAACAGCTACTCTCTCTCA 936
602 GlyAlaIaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPh 618
935 TCGCTACTCATTCATCACTCACTCACTCACTCACTCACTCACTCTGATT 886
618 e.....TyrAlaHisAspSerSerGlyLys... 626
885 CATGATCANACTAGCCTTTCGMACTCAACAGCACTAACCGGCTG 836
627ProIleAspAsnTrpHisArgSerLeuGlyTyrLeuPhe 640
835 TNAGCATAGTATAGNGANGCTMACTCGTGTATGTTATGTTGTCTTC 786
641 GlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIaG 657
785 TCTGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 743
657 lYglnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluThr 673
742TCTATCTCGTANTTTTACTACTCATACGACTACT 710
674 ThrSerTyrIleAlaThrValGlnAlaGlnIleuAlaThrSerLeuMet 690
709 CTNACCGCATTTTGTCTCNATCTCATCATATTATGACCGGTGTGATAC 660
690 sIleSerAlaGlnAlaCysTyrAsnGlu.....SerIleHisG 703
659 TGANCTGCTTCACACNCTCCACGACTCACACACACACAGCTCCACG 610
703 luleuLysThrLysTyrArgSerPheSerLys 713
609 CAGCATCATCAACATACACNAGATCAANA 578

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:25:51 ; Search time 27.13 Seconds
(without alignments)
3793.164 Million cell updates/sec

Title: US-09-677-752-4
Perfect score: 4533
Sequence: 1 MRPDHMFCCCAILSSTA.....CVLRGSHSYSLDGLTTRF 878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organellar:*
9: SP-phage:*
10: SP-plant:*
11: SP-protist:*
12: SP-unclassified:*
13: SP-vertebrate:*
14: SP-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4509	99.5	878	2	084882 chlamydia t
2	3543.5	78.2	867	2	09PL41 chlamydia m
3	947.5	20.9	928	2	092398 chlamydia p
4	919.5	20.3	930	2	09RB66 chlamydia p
5	915.5	20.2	930	2	092393 chlamydia p
6	896.5	19.8	926	2	P71135 chlamydomph
7	894.5	19.7	1276	2	09JRW2 chlamydia p
8	894	19.7	1407	2	092899 chlamydia p
9	883	19.5	928	2	09RB65 chlamydia p
10	853.5	18.8	914	2	086163 chlamydia p
11	845	18.6	928	2	086164 chlamydia p
12	845	18.6	949	2	09K299 chlamydia p
13	837.5	18.5	772	2	09RB71 chlamydia p
14	816.5	18.0	973	2	092896 chlamydia p
15	816.5	18.0	993	2	09K2A1 chlamydia p
16	790.5	17.4	846	2	P71133 chlamydomph
17	779	17.2	847	2	P71132 chlamydomph
18	776.5	17.1	839	2	P77792 chlamydomph
19	773	17.1	936	2	092898 chlamydia p

20	772	17.0	936	2	09J542 chlamydia p
21	758	16.7	987	2	09PL45 chlamydia m
22	758	16.7	1013	2	084879 chlamydia t
23	740	16.3	922	2	0929G5 chlamydia p
24	739	16.3	922	2	09K1Y9 chlamydia p
25	735	16.2	922	2	0924H9 chlamydia p
26	708.5	15.6	841	2	0923A1 chlamydia p
27	575.5	12.7	946	2	09RB60 chlamydia p
28	574.5	12.7	946	2	0928B0 chlamydia p
29	566	12.5	947	2	092813 chlamydia p
30	566	12.5	947	2	09JSE2 chlamydia p
31	562.5	12.4	649	2	P71134 chlamydomph
32	557.5	12.3	978	2	09RB63 chlamydia p
33	556.5	12.3	978	2	092895 chlamydia p
34	546.5	12.1	712	2	09RB73 chlamydia p
35	540.5	11.9	938	2	092883 chlamydia p
36	531	11.7	934	2	092882 chlamydia p
37	531	11.7	934	2	09JSE7 chlamydia p
38	531	11.7	952	2	09K2A5 chlamydia p
39	510.5	11.3	1520	2	09PLB0 chlamydia m
40	503.5	11.1	1609	2	09RB58 chlamydia p
41	503.5	11.1	1609	2	0926U5 chlamydia p
42	493.5	10.9	986	2	09PY33 chlamydia m
43	487	10.7	1016	2	084880 chlamydia t
44	484.5	10.7	975	2	084417 chlamydia t
45	482.5	10.6	976	2	09PL47 chlamydia m

ALIGNMENTS

RESULT 1
084882 PRELIMINARY: PRT: 878 AA.
AC 084882:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN I.
GN PMPI.
OS Chlamydia trachomatis.
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; Pubmed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AF001361; AAC68472.1;
SQ SEQUENCE 878 AA: 95592 MW; DFLFIA31707BE48B CRC64;

Query Match 99.5%; Score 4509; DB 2; Length 878;
Best Local Similarity 99.5%; Pred. No. 9,9e-313;
Matches 874; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MRPDHMFCCCAILSSTA	VFQDPLGFTALLTKNPNHYVCTFFDDCTMESLFPALCA	60
Db	1	MRPDHMFCCCAILSSTA	VFQDPLGFTALLTKNPNHYVCTFFDDCTMESLFPALCA	60
Qy	61	HASDDPLVYGNSYCWFSKLTHTDPEALFKKGLDSTONFRLSTOCSSKSSPSI	120	
Db	61	HASDDPLVYGNSYCWFSKLTHTDPEALFKKGLDSTONFRLSTOCSSKSSPSI	120	
Qy	121	THKNGQLSRNNSMFCRNHAEAGSGAISADAFSLQHNLYLFTAFENSSKNGAIOA	180	
Db	121	THKNGQLSRNNSMFCRNHAEAGSGAISADAFSLQHNLYLFTAFENSSKNGAIOA	180	

OY 840 FRITNORALHRLQTLNLNSCVLRGOSHSYSLDGLTTRF 878
 DB 829 FRITNORALHRLQTLNLNSCVLRGOSHSYSLDGLTTRF 867
 RESULT 3
 ID 092398 PRELIMINARY: PRT: 928 AA.
 AC 092398;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE OUTER MEMBRANE PROTEIN PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
 GN OMP10 OR PMP_9 OR CP0306.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=63358;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CWL029;
 RX MEDLINE=9206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RT Nat. Genet. 21:385-389(1999).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-VR1310;
 RA Hjeno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christensen G., Birkelund S.,
 RT "Chlamydia proteins containing the GAI-repeat belong to a subfamily
 RT of autoexporting pathogenicity factors.";
 RT Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-0138;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA.";
 RT Nucleic Acids Res. 28:2311-2314(2000).
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Ullrich T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39.";
 RT Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AE001628; AAD18591.1;
 DR EMBL: AJ133034; CAB37069.1;
 DR EMBL: AP002546; BAA98655.1;
 DR EMBL: AE002192; AAF38163.1;
 DR TIGR: CP0306;
 KW Signal;
 FT CHAIN 1 27 POTENTIAL.
 FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.
 SO SEQUENCE 928 AA: 98332 MW: 58910A8F04F12219 CRC64;

Query Match 20.98; Score 947.5; DB 2: Length 928;
 Best Local Similarity 31.68; Pred. No. 5.2e-59;
 Matches 288; Conservative 137; Mismatches 336; Indels 149; Gaps 32;
 OY 85 TDPREALFKKGLDLSNFRF-LSFTDCSSSESSPIIHQNG-QLSLRN----- 132
 DB 52 TMDGTYNLGDSVSTNAGSPALTLASCFKRTTGNLSFGSHGQVFLQNLIDANCTFT 111

OY 133 ----NGSMF-----CRNHAEGSGAISAADAFSLQHNLYLFTAFENSSKNGGAI 178
 DB 112 NTAANKLSFSGFSYSLIQTNTATTGAIKSTGACSTQSN-SCYFQONSNNDGAL 170
 OY 179 QAOTFSLRNVSPISFARNRADLNGAICCSNLCISGN-VNPLFTGNSATN----- 229
 DB 171 QSSISLSLNP-LTFARKKATOKGALYSTGTITLNTLNSASFENNTAANGAIYTE 229
 OY 230 -----GGAICISDLNTSEKSLSLACQETLFASNAK 265
 DB 230 ASSFISNKAISFINNSVATSAATGAIYCSS--TSAPKPVLLSDNGELNFGNTAITS 287
 OY 266 GGAIVAKHMYLRANGVVSFINNS-----AKIGAIYQSGSLIAGESYVFNONS-- 318
 DB 288 GGAIVYDNLVLSSGPTLRNNNSAIDTAAPLGAIALDSGSLSLAGDITFEGNTIV 347
 OY 319 -QRTSDGLVRNAIV-EKDAISLSLEARNGD-ILFEDPIVDESSKESPLPSLOASYT 375
 DB 348 KGASSSQTTRNSINIGNTNAKIVOLRASQNTIYFDPT-----TSITALLS 396
 OY 376 SPPPATASPLVIQTSANRSVIFSSERLSEBKT-PNULTSOLOQPIELASGLVLKRAV 434
 DB 397 DALNLNGPDLACNPAYOGTIVFSGEKLSEAEADNLKSTIQOPLTLAGGLSKGYT 456
 OY 435 LSAPLSDPQALLMEAGSLTSSDKLATLSIPHSIDTEKSYTIIAPNLSIQKIFL 494
 DB 457 LVAKSFQSQPSGLTMDAGTLETADGITNNLVLVNDSLEKTKKATLAKTQAS-QTVVL 515
 OY 495 SNS-----GDENFVENVELLSKEQNN--IPLTLSKEQ-SHLPLPDGNSL----- 537
 DB 516 SGSLSLYDPSGNYEDV-----SMNPQVFSCLTLTDPPANHIID--LAAPLEKNPI 568
 OY 538 HPGYQGMWTFMSKDSDEGSLIA--NWTPKNYVPHPEROSTLVANLMTYSDMAVQSM 595
 DB 569 HMGVQGMWALSMQEDPATKSKAATLTWTKTGYNPNPERGTLVANTLWGSFVDRSIOL 628
 OY 596 INTIANGALETGWSGAVSNLFYAHDSGKPIDNMHRLSLGLFTGSHSDHSCLA 655
 DB 629 VAKRVROSOETRGICWEGISNFF--HKDSRK INKGFHISAGYVAGTTLASDNLITAA 666
 OY 656 AGOLLKSSDSFITSFTETTYIATVQAQ-LAT-----SLMK-----ISAQACYN 698
 DB 687 FCOLFGKDRHFLINKNASKAYASLHQLALATLSSPSLTYLGSSESGQVLDQAQSYI 746
 OY 699 ESJHELTKYRSFSGEGSGMWSVAAGVGCASIPYVS-NGSLFSSFSIFSLQGEFGT 757
 DB 747 YSKMTKTYTQAQPK-GESSWYNGCALLESAPHLALSHGCFHAYFPFIVEASYIH 805
 OY 758 QDGFEE-SSGEISFSASSFRNISLPICITFEKKSKQKTRTYVFLGXYIDDKRDVESGP 816
 DB 806 QDSFKERNITLVKRFSDGDLINVSVPICITFEFRSNERASYEATVIVADVVRKKNPDC 865
 OY 817 VVLLKNAVSDAPANADNSRA-----YMFRLTNORALHRLQTLNLNSCVLRGOSHSY 868
 DB 866 TALLINNTSKKTGTGTLNSRQAGIGRAGIFTAFA-----PNLEVTNLSMETIGSSRSY 918
 OY 869 SLUDIGTTRF 878
 DB 919 NADGKGFQF 928
 RESULT 4
 ID 09RB66 PRELIMINARY: PRT: 930 AA.
 AC 09RB66;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
 GN PMP_8 OR CP0307.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

```

Db 349 NTLTSTASTRNAIYLGCSSAKITNLRAAOCOSIYFYDPIASNTTGASDVL----- 400
Oy 375 TSPPTASTPLVYIOTISANSVIFSSERLS-EEKTPDNLTSLQOPIELKSGVLKDBA 433
Db 401 TTNQPDNSPL-----DYSGTIVFSGEKLSDAKKADNTSILKOPLALASGLTALKGVN 456
Oy 434 VLSAPLSDDPOLLIIMEAGTSLKTSDD-LKLTATLSIPLHSLDTEKSVTIHAD-----N 486
Db 457 ELDVNFTQTEGSTLLMQGTIKKADTEAISLKLIVYDLSALEGNKSVSIETGANKTIT 516
Oy 487 LSTQKIFLNSGDENFYENVELLSKEONNIPLLITLSKEOSHLPDGNLSS-----HF 539
Db 517 LNSPLVFODSSG--NFYESHITINOAFTOPLVYFTAATAASDIYI-DALLTSPVQTEPHY 573
Oy 540 GVOGDPTSMKSDDEGHSLIANMTPKNYVPHPEROSTLVANTLTMTNTYSDMQAVOSMINTI 599
Db 574 GTGGHBEATADTSTAKSGMTMTVTGYNPNPERRASVYFDSLMASTDTRTLOQIMTSQ 633
Oy 600 AHGAVLFGTWSAVSNLFYAHDSSGKRPIDNMHRSLSGLYLGISTHSLDHSFCLAGOL 659
Db 634 ANSIYQORGLMAGSTANFPH-KDKSGTN-QAFPHKSYGYIVGSAEDFSENIFSVAFCOL 691
Oy 660 LKGSOSFITSTETSTIATVQAQ-----LATSLMK-----ISAQACINES 700
Db 692 FCKDKDLFIVENTSHNYLASLYLQHRAFLGGLPMPSPGSIITDMLKDLPLILNQLSYST 751
Oy 701 IHLKTKYRSFSEKESGSMHSAVSGEVCASIP1-VSNGSLFSSFSIFSKLOGFSGTOD 759
Db 752 KNDMDRTYTSY-PEAQOSMTNNGSGALEGLSLYLLEKPEFPGYIFPLKFAVYRQ 810
Oy 760 GFEESGSEIRSFSSASFRNISPILGITEFKKSOKTRTYVYFLGAYIIDLKRDVSGCVL 819
Db 811 NFVESGAEANAFDDGDLVNCISIPGILKISEDEKKNFEISLAYIGDYVRKNPERSRSL 870
Oy 820 LKNAVSDAPMANIDSRAYFRLLTNORALH-RLQTLNVCVLRGSHS5LDLGTYYR 878
Db 871 MVSQASWTSJLCKNLARQAFLASAGSHLTLSPHVELSGEAAYELNGSAHY1NVDCGLRYSP 930

RESULT 6
P71135 PRELIMINARY: PRT: 926 AA.
ID P71135;
AC P71135;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE PUTATIVE 98 KDA OUTER MEMBRANE PROTEIN.
OS Chlamydia pneumoniae.
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVINE ABORTION S26/3;
RA Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U72499; AAB1818.1;
SQ SEQUENCE 926 AA: 96439 MW: 3E75E5EF594750F CRC64:

Query Match 19.8%; Score 896.5; DB 2; Length 926;
Best Local Similarity 30.3%; Pred. No. 2,2e-55;
Matches 257; Conservative 145; Mismatches 341; Indels 105; Gaps 25;

Oy 98 LSTONFRLLSTDDSSKSSPSIIH-----QKNGQLSLRNGSGSPFRNIAEGSGAIS 151
Db 117 LTLTDFSKLSFKRCPD-----SLVNTGKAMKSGGALLNANNASILFDQNTSAENGGAIS 171
Oy 152 ADAPSLQNYLFTFAEENSXKNGCAIOQTFS-LSRVNPSISFARNADLNGAICG-S 209
Db 172 KKAFTSLSSSEISFTTTSTAKKGAIAATGTAHLSNDQGTFRFGCNTAVNSGGAIVSEA 231
Oy 210 NLICSGNVNPLFTGNSATN-----GGAICCSIDLNTSEKSGSLACQOETLFASSNAKE 264

```

```

Db 232 SMTIAGN-NHVAFSNNAVSGSSDCCGCAIHC---SKTGSAPLITITDNKVLFEENTSSA 287
Oy 265 KGAIAAKHMYLYNGPVSEFINN---SAKIGGAIAIOSGSLISLAGESVLPQNNR 320
Db 288 KGAIAITDKLILISGPTAFINNKVTHATPKGAGIAGIANGECSLTAEHGDIITFQNNLMA 347
Oy 321 TSDQGLV-RNATYLEKDALISLEARN-DILFDDIVDESSSEKSPFLSSQASTSTPT 378
Db 348 TQDNATIKRINAIIEGNGKRVNLAASGKTISEYDPIVEGNAAADLLINKAEGKTT--- 404
Oy 379 PATASPLVYIOTSANRSVIFSSERLSEEE-KTPNLTSLQOPIELKSGVLKDBA 437
Db 405 -----YNGRIITFSGEKLTEQAAVADNLKTTPTPTTLAAGELVLRSGVEVA 452
Oy 438 PSLSDPOLLIIMEAGTSLKT-SSDKLATLSIPLHSLDTEKSVTIHA-----PNLSIQKI 492
Db 453 KTVYQTAGSLILMDAGTKLSAKTEDATLTNLAINPTLDGKRFVAVDAVAGKNTLSGA 512
Oy 493 FLNSGDENFYEN-----VELLSKEONNIPLLITLSKEOSHLPDGNLSSHFGY 541
Db 513 IGYIDPTGKFEYENHKLNDTLALGIGLSGKS-----VTTTNPVSHV--VGVAETHYGY 564
Oy 542 QGDMTESW-KDSD---EGHSLIANMTPKNYVPHPEROSTLVANTLTMTNTYSDMQAVOSM-- 595
Db 565 QGMWVSVMKADNNSDKRTQTAIFTWKTCYVNPERRAPVLSNLSGFTDLASIQDYLE 624
Oy 596 -INTIAGGAVLFGTWSAVSNLFYAHDSSGKRPIDNMHRSLSGLYLGISTHSLDHSFC 653
Db 625 RVSYSLETRR---GLWVSGIGNF--HKDRNANERKFRHISGGYVLAGATTNTSRDLSLS 679
Oy 654 LAAGQLKGSOSDFSTETSTIATVQAOLATSL-----MKI 691
Db 680 VAFQQLFAKDKDYLKSNKAAVYAGVYQHSFDDLTBLFNGPNTCGSGFKEPIPL 739
Oy 692 SAQACYNESIHETKTKYRSFSEKESGSMHSAVSGEVCASIP1VYNSGSLFSSFSIFSKL 751
Db 740 DAQITTYCHTANMTTSTDY-PEYKSGMGNLTLGLTSTVPIPVSSSIFDSTAPPAKL 798
Oy 752 QGESGTOGFEESGSEIRSFSSASFRNISPILGITEFKKSOKTRTYVYFLGAYIIDLKRD 811
Db 799 QVYVAHDDDEKPTTEGRVSESSDLNVSPDIGIKFEKLSYGERSAVDLTLMYIPQYRH 858
Oy 812 VESGPVVLKNAVSDAPMANIDSRAYFRLLTNORALH-RLQTLNVCVLRGSHS5YSL 870
Db 859 NPSGMTALINDVSWLTATNLARQAFIVRAGNHIALTSGVEMFSOPFELRSSRMYNV 918
Oy 871 DLGTYRF 878
Db 919 DLGAKVAF 926

RESULT 7
O9JRW2 PRELIMINARY: PRT: 1276 AA.
ID O9JRW2;
AC O9JRW2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G/I FAMILY (POLYMORPHIC MEMBRANE
PROTEIN G FAMILY).
GN PMP_6 OR CP0309.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RA MEDLINE=20330349; PubMed=10871362;
RA Shira M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CW029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).

```

QY 521 LSKESQSHLHPDGN-----LSHFEGYQDWTFSWKDSEGH-SLIANTPTKNVYHPER 573
 DB 1024 LSTSGTVDLPDGNFIPSSMAAPDYQSGSMTLVKPVKAGKVTLLVAEQALCYTPKPEL 1083
 QY 574 QSTLVANTLVNTYSDMQAVQSMINTIAHGAVLFGTWGSAVSLFPAHDSGCPIDNMHH 633
 DB 1084 RALVFNLSMNVNINHISIOEITATMSDAPSHPIGIMIGIGNAF--HODKOKENAGFRL 1141
 QY 634 RSLGYFGISTSLDHSFCLAGOLLGKSSDSFISTSTETTYTATVQAO-----LATS 687
 DB 1142 ISRGYIVGSSMTTPOETRAVAFSOLFSGSKDYVSDIKSQYAGSLCAOSSVITLHSS 1201
 QY 688 LMK-----ISNAQYNESIHELKTRYSKEGSGHSAVSEVCA 730
 DB 1202 LRRHVLKVLPELPGETPLVLHQVSYGRNHNMTTKLAN-NTQKSDMDSHFAVEVG 1260
 QY 731 SLPIVNSGSLFSSSIFSKLQSGTQDGFESSGEIRFSASSFRNISLPIGITEKK 790
 DB 1261 SLPVDLNTYRL-TSYSPRYKLVQVSVNOKGFOEVADPRIFDASHLVNVSIPMGLTFKHE 1319
 QY 791 SOKRTYVFLAGYIQLDRVDSGPVLLKNVSWDAPMANLDSRAYMRLTNO--RALH 849
 DB 1320 SAKPPSALLTLGAYAVDAYRD-HPHCLTSLTNGTSMSTFATNLSTROFPAEASGHLKLH 1378
 QY 850 RLQTLNVSCVLRGQSHSYSLDGLTYRF 878
 DB 1379 GLDCFASGCELRSSRSRYNANGCTRYSF 1407
 RESULT 9
 Q9RB65 PRELIMINARY: PRT; 928 AA.
 ID Q9RB65: Q9RB64: Q9SBP2:
 AC 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
 DE PRECURSOR) (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
 GN PMP.10 OR OMP5 OR CP0303.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83358;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-J138:
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RN Nucleic Acids Res. 28:2311-2314(2000).
 RP SEQUENCE FROM N.A.
 RC STRAIN-VRI1310:
 RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
 RA Christensen G., Birke Lund S.,
 RT "Chlamydia proteins containing the GCAT-repeat belong to a subfamily
 RT of autoexporting pathogenicity factors."
 RN Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39:
 MEDLINE=20150255; Pubmed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Unayam L.A., Uterback T.,
 RA Berry K., Bess S., Linher K., Weidman J., Khoriti H., Colaven B.,
 RA Bowen C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonen J.,
 RA McClarity G., Salzberg S.L., Eisen J., Fraser C.M.,
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
 RT pneumoniae AR39."
 RN Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL: AP002346; BA98657.1;
 DR EMBL: AJ133034; CAB37071.1; -

DR EMBL: AE002192; AAF38160.1; -
 DR TIGR: CP0303; -
 KW Signal.
 SQ SEQUENCE 928 AA: 97229 MW: 059005206A1DD0E1 CRC64:
 Query Match 19.5% Score 883; DB 2; Length 928;
 Best Local Similarity 30.6% Pred. No. 2.1e-54;
 Matches 288; Conservative 145; Mismatches 401; Indels 108; Gaps 31;
 QY 12 CAALISTAVLFCGDDPL--GETALLTKPNHNYVC-----TFEEDCTMESTL--FPALCAHAS 63
 DB 20 CSTVFATIAENICPSDFDSTNTGTYTPKNTTGTIDYITLIGITLQNLDSALAKGCF 79
 QY 64 QD--DPLVYLGNSYCWVFSKRLHTTDPKEALFKKGLDSIONFRLSFTDCSSKE--SSPS 119
 DB 80 SDTTELSIFAGKGY--LSPLNIKSSAEG-----AALSVTTDKNLSTGFSLLTFAAS 132
 QY 120 IT-----HOKNQLSLRNNGSMSPFCNNHAEAGGGAISADAFSLQHYLFTAEENS 170
 DB 133 SVITTFSGKCAVCGCDLTFDNNGTILFKODYCEENGALSTNLKSTGSISEGNK 192
 QY 171 SKG---NGCAIOAQ--TPESLRNVSPISFARNRADLNGAI--CCSNLCSGNVPLPFTGN 225
 DB 193 SSATGKKGAIATGTVDTINNTAPTLFSNNIAEAGAGAINSTGCTTGNTS-LVFSEN 251
 QY 226 SAT-----NGCAICISIDLNTSEKGSLSLACNDETLFASNSAKEKGAIVAKHMYLR--N 279
 DB 252 SVATAGNGAL-----SGDADVTISGNQSVTFSGNQAVANGGAIYAKKLTLASGG 303
 QY 280 GPVSFLNN-----SAKIGAIQISGSGSLILAGEGVLFONNS--ORTSDGLVRNHYL 333
 DB 304 GGISFSNNIVQGTAGNGAISTILAGECSLSAEAGITFNGAIAVATPQTTRNSIDI 363
 QY 334 EKALISLSEARNG-DILFPDPIVOESSKESPLPSLSLQASVSPATASPLVIOFSAN 392
 DB 364 GSTAKTITNLRAISGHSTIFFYDPTTANTASTDITLNNKADAGN-----STDYS 412
 QY 393 RSVLFSSEIRLSEEE-KTPDNLTSQLQPIELKSGRLVLMKRAVLSABSLSQDQALLME 451
 DB 413 GSIVFSGERLSEDEAKVADNLSTLKPVTLTAGNLVLKRGVTLDTGFTQAGSSVIMD 472
 QY 452 AGTSLKTS--DLKLATLSTLHSLDTEKSVTHAP-----NLSIQKIFLSNGDEMYE 504
 DB 473 AGTTLKASTEEVTLTGTSIPVDSLGECKKVYIAASAASKNVALSGPILLDNOG--NAYE 530
 QY 505 NVLELSKEON-----NPLPLLSKQSHLHPDGNLSHFEGYQDWTFSWKD-----SDEGH 556
 DB 531 NHD-LGKTQDFSEVOLALGTATTDDVPVAPVATPTHTYGTGWTGNTWDDTASTPKTK 589
 QY 557 SLIANTPKNYVPHPEROSTLVANTLVNTYSDMQAVQSMINTIAHGAVLFGTWGSAVSN 616
 DB 590 TATLAWTNTGILPNEROGLVLPNSLWGSFSDIQAIGYIERSALITLCSDRGFMAAVAN 649
 QY 617 LEYADSSGKPIIDMHHRSGLYFGISTHLDHSFCLAGOLLGKSSDSFISTETTSY 676
 DB 650 -FLDKDKKCKE-KRYRHRKSGGAYGAAGTCSENLISFACQDLFGSKDPDLVAKNHTDY 707
 QY 677 IATVQAOALATSL-----MKISAQYNESIHELKTRYSKEGSGFG 717
 DB 708 AGAFYIOHITCEGFGICCLLDKLPKSHKPLVEGGLAYSHVSNDKITAT--PEYKG 766
 QY 718 SMHSVAVSGEVCAISPIVNSGSLFSSSIFSKLQSGTQDGFESSGEIRFSASSFR 777
 DB 767 SMCNNAFNMWMLGASHSYPEVLCFDTYAPYIKLNTLYIHODPSSEKGTGRSFDOSNL 826
 QY 778 NISLPIITTEKSKQKRTTYTYFFGAIYIQLDRVDSGPVLLKNVSWDAPMANLDSRA 837
 DB 827 NLSLPIVCKFEKESDCNDPFDLTLSVYDPIRNDPKCTTALVIGSASWETVANNLARO 886
 QY 838 YMFRLTNRALHRL-OTLVNVSCVLRGQSHSYSLDGLTYRF 878
 DB 887 LQVAGSHYAFSPMFEVLGQFVEFVRGSSRLYINVDLGGKTOF 928

KW Signal 1 17 POTENTIAL.
FT SIGNAL 18 928 OUTER MEMBRANE PROTEIN 4.
CHAIN 18 928
SEQUENCE 928 AA: 98903 MW: 788BCDD62C911402 CRC64;

Query Match 18.6% Score 845; DB 2; Length 928;
Best Local Similarity 28.8%; Pred. No. 1e-51;
Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;

97 DLSDIQRFLSTFDCSSKESPPS-IIHOKNGOLS-----LRNNGSMFPCNNHAGSGCA 149
113 NLTFSGSLISF-----DSSPTVTYTGQTLSSAGGVNLEIRKLVAAGNFSTADGCA 166
150 ISADAFSLQHNLYLFTAFEPENSSKNGGALIQQTFS-LSRNVSPISFARNRDLNGALIC 208
167 IKGASFLITGSGDALFNSNSSSTKGALATATAGARIANNNGYRFLSNIASTSGAID 226
209 SNLICSGVNPDLFTTNSA-TNGAICISDLNTSEKGS--LSLACNOETLPASNAKEK 265
227 EGTSLISNNKFLYEPGNNAKTGGAIC-----NTKASGSPELIISNNKTLIFASVVAETS 281
266 GGAIYAKHMYLRNGPVSFINNSAKI-----GGAIYIAGSGSLISLAGSGVLFQNSQRT 321
282 GGAIAHAKKLALSSGGFTFLRNWSSATPKGALISIDAGSELISAETGNTIFVARNITLT 341
322 --SDOGLVRNATYLEKDALISLSE-ARNGDILFDPPIVOESSKESPLPSSLOASVTSPT 378
342 TGSITDTPRRNAINIGSNCKFTELRAKAKNHTIFFYDPTISEGSSDVLEKINNGSAGALNPY 401
379 PATASPLVYIOTSANRSVIFSSERLSEEE-KTPDNLTSQLQPIELKSGRLVAKRAVISA 437
402 QGT-----ILFSGETLTADELKADNLSKSTFQVSLSGKLLQKCVTL 448
438 PSLSDPPALLIMEAGTSLKTS--DLKATLISPLHSIDTEKSVIIRHNPUSIOKIFLSN 496
449 TFSQEAQSLGMDSGTTLTAGSITITNIGINVDLSGLKQVSLAKKAS-NKVIYSG 507
497 -----SGDENFEYENVELLSKEO-----NNIPLLTSLKESQSHLHPGNLSS 537
508 KLNLDIDEGNIYES-HMFSHDLFLSLKTIYDADVDNTVDISL-----IYVPAEDPNS 560
538 HFCYOGDMTFFSKMDSDEGHS--LIANMTPKNVYVPEROSTLVANTLWNTYSMDQAVOSM 595
561 EYFGQGMVNMVNTDTATNTEKATATMTKTGFVSPERKSALVCLNTLWCVFIDISLOOL 620
596 INTIAHGCAVLEFGTWSAVSNLFYAHDSGKPIDMNMHRSGLYLGISTHSLDHSFCLA 655
621 VEIGATGMEHKQGFVWSSMTN--FLHKTGDENRKGFRHTSGGVYIGSAHTPKDLETF 678
656 AGOLLKSSDSFTSTETSYIATV-----QAOLATSLMK----- 690
679 FCHLARDKDCGAIANNSTGTGTFEFKHSHTLOPQNTYLRGRAKFSEAIKFPREIPL 738
691 -ISAQACYNESIHLELTKYRSEKGFSGWSHSAVAGEVCASIP-IVSNGSLFSSSIF 748
729 ALDVOVFSHSDNRMEHTYSL-PESEGSMSNECIAGIGLDLPFLSNPHLFTFTFIPO 797
749 SKLQFGSGIDQDESSSGIRSFASAFRNISLPIGITEKKSOKTRITYYFLGAYIOOL 808
798 MKVENYVYSONSFSSSDGRGFSIGRLNLISIYVGAKVQODIGDSYTYDLISGFFVSIV 857
809 KRDVSGPVLILKNAVMDPAMNLDRAVYFRLTNQALHRLQTLNLSVC-----V 860
858 YANNOSTITVLVMSDPMIRGNLSROAFLRGSN-----NYYNNSCELFGHYAME 910
861 LRQSHSYSLDLGTYRF 878
911 LRSSSNYVNDVGTKLRF 928

ID 09K299 PRELIMINARY; PRT: 949 AA.
AC 09K299:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0302.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_Taxid=83558;
RN (1)
RP SEQUENCE FROM N.A.
RC MEDLINE=20150255; Pubmed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umeyam L.A., Ullrichback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AEO02192; AAF38159.1; .
DR TIGR: CP0302; .
SO SEQUENCE 949 AA: 101357 MW: A00B09E16C699BE3 CRC64;

Query Match 18.6% Score 845; DB 2; Length 949;
Best Local Similarity 28.8%; Pred. No. 1.e-51;
Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;

97 DLSDIQRFLSTFDCSSKESPPS-IIHOKNGOLS-----LRNNGSMFPCNNHAGSGCA 149
134 NLTFSGSLISF-----DSSPTVTYTGQTLSSAGGVNLEIRKLVAAGNFSTADGCA 187
150 ISADAFSLQHNLYLFTAFEPENSSKNGGALIQQTFS-LSRNVSPISFARNRDLNGALIC 208
188 IKGASFLITGSGDALFNSNSSSTKGALATATAGARIANNNGYRFLSNIASTSGAID 247
209 SNLICSGVNPDLFTTNSA-TNGAICISDLNTSEKGS--LSLACNOETLPASNAKEK 265
248 EGTSLISNNKFLYEPGNNAKTGGAIC-----NTKASGSPELIISNNKTLIFASVVAETS 302
266 GGAIYAKHMYLRNGPVSFINNSAKI-----GGAIYIAGSGSLISLAGSGVLFQNSQRT 321
303 GGAIAHAKKLALSSGGFTFLRNWSSATPKGALISIDAGSELISAETGNTIFVARNITLT 362
322 --SDOGLVRNATYLEKDALISLSE-ARNGDILFDPPIVOESSKESPLPSSLOASVTSPT 378
363 TGSITDTPRRNAINIGSNCKFTELRAKAKNHTIFFYDPTISEGSSDVLEKINNGSAGALNPY 422
379 PATASPLVYIOTSANRSVIFSSERLSEEE-KTPDNLTSQLQPIELKSGRLVAKRAVISA 437
423 QGT-----ILFSGETLTADELKADNLSKSTFQVSLSGKLLQKCVTL 469
438 PSLSDPPALLIMEAGTSLKTS--DLKATLISPLHSIDTEKSVIIRHNPUSIOKIFLSN 496
470 TFSQEAQSLGMDSGTTLTAGSITITNIGINVDLSGLKQVSLAKKAS-NKVIYSG 528
497 -----SGDENFEYENVELLSKEO-----NNIPLLTSLKESQSHLHPGNLSS 537
529 KLNLDIDEGNIYES-HMFSHDLFLSLKTIYDADVDNTVDISL-----IYVPAEDPNS 581
538 HFCYOGDMTFFSKMDSDEGHS--LIANMTPKNVYVPEROSTLVANTLWNTYSMDQAVOSM 595
582 EYFGQGMVNMVNTDTATNTEKATATMTKTGFVSPERKSALVCLNTLWCVFIDISLOOL 641
596 INTIAHGCAVLEFGTWSAVSNLFYAHDSGKPIDMNMHRSGLYLGISTHSLDHSFCLA 655
642 VEIGATGMEHKQGFVWSSMTN--FLHKTGDENRKGFRHTSGGVYIGSAHTPKDLETF 699
656 AGOLLKSSDSFTSTETSYIATV-----QAOLATSLMK----- 690

92 FK-----EKGDLSIQNFRFLFTDCSSKESPS-----IHK-----NGQLSLRNNGMS 137
109 SSVITONPELCPISFSGFSOMIFDNCESLTDTSASNVIIHSAIYATTPMLFNNISIL 168
138 FCNHAEGSGAISADAFSLQHNLYLFAFEENSSKNGCAIQ-AQTSLSRNVSPISFAR 196
169 FOYNSAGFGAIAIRGTSITIENTKSLFNGNCSISNGALTGSAIINLINSAPVIFST 228
197 NRADLNGAICCS--NLICSGNVNPLFTGNSATNGAICISDLNTSEK-----245
229 NATGIYGAATYLTGSMLTSGNLGSLVFNNSKSGAIYANGVNTSNNSDLTFOUNTA 288
246 -----SLSLACNOETLFAASNAKE 264
289 SPONSLPAPTPPPPPATVPLTGGAIFCTPPATPPPTGVSLSLTISGENSVTFLENIASE 348
265 KGAIAVAKHMYLRVNGPVSFINSKIGAIATIOSGSLSLTAGESVLPONNSORTSDQ 324
349 OGGAALYKIKISIDENKSTIFLGNTAGKGAIAIPESGELSLSANQDILFNKNLSITSGT 408
325 GLVRNAIYLEKDALISLEARNG-DILFDPPIVOESSKESPLPSLQASVSTPATAS 383
409 P-TRNSIHFGKDAKATLGATOGYTLFYDPITSDLSAAS-----AAA 451
384 PLYIOTSANR-----SVIFSEERL-SEEEKTPDNLTSOLOQPIELKSGRLVKDRAVLS 436
452 TVVNPKNASADAGVSGTIVSGETLTATEATPANATSTLNKLEEGTLALNGATLN 511
437 APSLSDOPALLIMEAGTSLKTS-----LKLATLSIPLHSLDTEKSVTHAPN--- 486
512 VHNTOPEKSVIYMDAGTTLATNGANNFCAITLNLKLVIMLSDIGTKAAVNVQSTNG 571
487 -LSIQIFLNSGDNFENVELLSKEONNIPLLTSLKQSHLHPD-----GNLSHF 539
572 ALTISGTLGLVKNSDCCDNHGMFNKDLQOVPLELKATSNVTVTDTSLCTNGIQOQSPY 631
540 GYGDWTFMSKDSDEGHSILANMTPKNVPHPEROSTLVANTLMNTYSDMOAVOSMINTI 599
632 GYOGTWEITITTT--HTVTGMKKTKGYLPHPERLAPLIPSLMANVIDLRAVSA--SA 687
600 AHG-----GAYLFGTWGSAVSNLFYA-HDSGKPIDNMHRSGLYLFGISLSDHSEFL 654
688 ADGEVDPKQSLIT--GITNEFHANHTGDAR--SYRHMGGYLLINTYTRITPPAASL 741
655 AAGOLLKSSDSFISTSTETSYIATVOALATSLMKIS-----AOCYNESIHELTK 707
742 GFGQLFTKSKDYLVGCHGSHNVYFATVYSNITKSLFGSSRFSGTSRVYTSKSNKVKTS 801
708 YRSFSEKFGSHSAVAGEVCAISPIVNSGSL-FSSFSIFSKIQGFSGTDDGEESG 766
802 YTKLPK-GRCSHNNCMIGELEGNIPILSSRILNLKOLIPVKADEVAYATNGIQENTP 860
767 EIRSFSSAFRNISLPIGITEKKSQKTRTYVYFLGAYIODIKRDESGPVVLLKANVAS 826
861 EGRIGHGHLNVAVPVGRCGNKSHNRPDYTIIVAYAPDVYRHNPOCDTTLPIGATW 920
827 DAPMANLDSRAVMTLNOALHR-LOTLLNVCYLRGOSHYSIDLGTTFYF 878
921 TSGNNLTRSTLLVQASSHTSYNDVLELFGHCGCDIRTSROYTLDIGSKLRF 973

RESULT 15
09K2A1 PRELIMINARY: PRT: 995 AA.

AC 09K2A1: 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0299.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

NCBI_TaxID=83558;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-AR39:
RX MEDLINE-20150255; Pubmed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002191; AAF38156.1; -;
DR TIGR: CP0299; -;
SQ SEQUENCE 995 AA; 105599 MW; 8562730FA8B683F CRC64;

Query Match 18.08; Score 816.5; DB 2; Length 995;
Best Local Similarity 26.44; Pred. No. 1,2e-49;
Matches 252; Conservative 158; Mismatches 404; Indels 139; Gaps 25;

37 NPNHVCFFFECDMESLFPALCAHAS-----ODDPLVILGNSYCMFVSKLHTDPREAL 91
71 DPKCTLCIFSGDLVIANLDMNAISRTSSCSFNRAGALQILKGGVFFLNIRSSADCAAI 130
92 FK-----EKGDLSIQNFRFLFTDCSSKESPS-----IHK-----NGQLSLRNNGMS 137
131 SSVITONPELCPISFSGFSOMIFDNCESLTDTSASNVIIHSAIYATTPMLFNNISIL 190
138 FCNHAEGSGAISADAFSLQHNLYLFAFEENSSKNGCAIQ-AQTSLSRNVSPISFAR 196
191 FOYNSAGFGAIAIRGTSITIENTKSLFNGNCSISNGALTGSAIINLINSAPVIFST 250
197 NRADLNGAICCS--NLICSGNVNPLFTGNSATNGAICISDLNTSEK-----245
251 NATGIYGAATYLTGSMLTSGNLGSLVFNNSKSGAIYANGVNTSNNSDLTFOUNTA 310
246 -----SLSLACNOETLFAASNAKE 264
311 SPONSLPAPTPPPPPATVPLTGGAIFCTPPATPPPTGVSLSLTISGENSVTFLENIASE 370
265 KGAIAVAKHMYLRVNGPVSFINSKIGAIATIOSGSLSLTAGESVLPONNSORTSDQ 324
371 OGGAALYKIKISIDENKSTIFLGNTAGKGAIAIPESGELSLSANQDILFNKNLSITSGT 430
325 GLVRNAIYLEKDALISLEARNG-DILFDPPIVOESSKESPLPSLQASVSTPATAS 383
431 P-TRNSIHFGKDAKATLGATOGYTLFYDPITSDLSAAS-----AAA 473
384 PLYIOTSANR-----SVIFSEERL-SEEEKTPDNLTSOLOQPIELKSGRLVKDRAVLS 436
474 TVVNPKNASADAGVSGTIVSGETLTATEATPANATSTLNKLEEGTLALNGATLN 533
437 APSLSDOPALLIMEAGTSLKTS-----LKLATLSIPLHSLDTEKSVTHAPN--- 486
534 VHNTOPEKSVIYMDAGTTLATNGANNFCAITLNLKLVIMLSDIGTKAAVNVQSTNG 593
487 -LSIQIFLNSGDNFENVELLSKEONNIPLLTSLKQSHLHPD-----GNLSHF 539
594 ALTISGTLGLVKNSDCCDNHGMFNKDLQOVPLELKATSNVTVTDTSLCTNGIQOQSPY 653
540 GYGDWTFMSKDSDEGHSILANMTPKNVPHPEROSTLVANTLMNTYSDMOAVOSMINTI 599
654 GYOGTWEITITTT--HTVTGMKKTKGYLPHPERLAPLIPSLMANVIDLRAVSA--SA 709
600 AHG-----GAYLFGTWGSAVSNLFYA-HDSGKPIDNMHRSGLYLFGISLSDHSEFL 654
710 ADGEVDPKQSLIT--GITNEFHANHTGDAR--SYRHMGGYLLINTYTRITPPAASL 763
655 AAGOLLKSSDSFISTSTETSYIATVOALATSLMKIS-----AOCYNESIHELTK 707
764 GFGQLFTKSKDYLVGCHGSHNVYFATVYSNITKSLFGSSRFSGTSRVYTSKSNKVKTS 823

DR TIGR: CP0308; -
 SQ SEQUENCE 936 AA: 100105 MW: 3981DB3C950AF95A CRC64:

Query Match 17.0% Score 772: DB 2: Length 936:
 Best Local Similarity 28.0% Pred. No. 1.7e+46:
 Matches 236: Conservative 143: Mismatches 378: Indels 86: Gaps 22:

97 DLSTQNRFLPSFDSCSKESPSS--IIHKNQCLSLNNNGSMFCRNHAGSGAISADA 154
 119 NLFPNDFSRISITISCBLLSPFGOCALKVGNLSLNGNQLIIFTONFSSDNGCVINTKN 178
 155 FSLQHNLYLFAFEENS-KNGGAIOAQFTSLRNVP--ISFARNRADNGAICCSNL 211
 179 FLISGTSQFASFRNQFTKCGGVVATGTITENSPGIVSFQNLAKSGGALYSTD- 237
 212 ICSGNVW-PLFPFGNSA-----TNGAICISDLNTESEKSLIACQETLIFASNSAKEK 265
 238 NCSITDFQVIFPDGNSAMEAQAQGAIC----TTTDK-TVLTGKNLSFTINITALTY 292
 266 GGAIVAKHMYLRNGVPSFINNSAKI-----GGAIIQSGSLIAGESVLFQNNNQ 319
 293 GGAISGLKAVISAGGPTLFGSNTSGSAGCGGCAIMIASAGELALATSQDITF-NNNQ 351
 320 RTSDQGLVRAIYLEKDAIISLEARNG-DILFDPIVOESSKESPLPSLOASTVTSPT 378
 352 VTNGSTSTRAINIIDAKVTISRAATGQSIYFDPITNPGTAASTDTLNINLADANS-- 409
 379 PATASPLVITOTSANRSVIFSESRSEBEKT-PDNLTSQLQOPIFLKGRVLKRAVLSA 437
 410 -----EIEYGAIVFSGEKLSPTEKAIANVTSTIRQPAVLAGDVLFGVTVTF 460
 438 PSLSDPQALLIMAGSLMT-SSDLKATLPLSLHSDTEKSVTHA---PULSTOKI 492
 461 KDLQSGSRLMDGTTLSKAKENLSLNGIAGVLSLIDGKNKALITEADKNISLSGT 520
 493 FLNSGDEFENVELLSKEONNIPLLTLESQSHLPDGNLS-----SHFGYOGD 544
 521 IALIDTSGSFYENHN--KSASTYRPLELTLYAGANGTITGALSTLQEPETHYGYOGN 578
 545 WTFSWKSDECHSLIANTPKNYVPHPEROSTLVANILMNTYSQMAVQSMINTIANHGA 604
 579 WQLSMANATSSKIGSINWTRTGYIPSPERKSNLPLNSLMGNFIDIRISINOLETKSSGP 638
 605 YLFTWGSVAENLFYANDSSGKPIDNMHRSLGYLFGISTSLDHSFLACQOLGKSS 664
 639 FERELMISGLANFFYRDSMPTR--HGFNHSIGYALGITATTAPADQLTFAFCOLFAADR 696
 665 DSFTSTETTYIATVQQLATSLMKIS-----AQACYN 698
 697 NHITGKNHGDYGAFLYHTEGLFDIANFLMGKATRAVPLSEISQIILPSFAKFSYL 756
 699 ESTHELKTY--RSPSKSGFGSMHSAVSGEVCASIPYVSGSLFSSFIETSLQSGSG 756
 757 HTDHMKTKYTYDINSIIK--GSMRNDACADLGASLPFVIVPYLKEVEPVKQYIYA 813
 757 TDQFEESGCEIRSPSASFRNISLPITFEKKSQKTRTYFYFGAIIODLKRDVESGP 816
 814 HQDPFEYTAGRAFKSELINVEIPIGVTFERDSKSKGTVDLMLKILDAYRRNPKCO 873
 817 VLLKNAVMDAPMANLDRAVMEFLTNQALH-RLQTLVANSVCLRGOSHSYSLDICTT 875
 874 TSLASDANMAYGTNLAROGFSVRAHNFQVNPHEIFGQFAFEVRSSSRNYTNLCSK 933
 876 YRF 878
 934 FCF 936

RESULT 21
 09PL45
 ID 09PL45 PRELIMINARY: PRT: 987 AA.
 AC 09PL45:

DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
 GN TC0263.
 OS Chlamydia muridarum.
 OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
 NCBI-TaxID=83560;
 RN
 RX SEQUENCE FROM N.A.
 RC STRAIN=MOPN / NIGG.
 RM MEDLINE=20150255: Pubmed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.,
 RA "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN
 RP SEQUENCE FROM N.A.
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
 RA Hickey E.K., Peterson J., Uterback T., Berry K.,
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
 RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,
 RA Salzberg S.L., Eisen J., Fraser C.M.,
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO02293: AAF39132.1;
 DR TIGR: TC0263;
 SQ SEQUENCE 987 AA: 104867 MW: 32079BD6BEB2DA42 CRC64:

Query Match 16.7% Score 758: DB 2: Length 987:
 Best Local Similarity 26.7% Pred. No. 1.8e+45:
 Matches 246: Conservative 125: Mismatches 353: Indels 198: Gaps 28:

100 IONRFLSTFDCSSKESPSTIIHKNQ-----QLSRNNGSMFCRNHAE 144
 121 IEAFDELILNMCN--SLVSVVPGTGTTSVPNSGRTYRTDVLIRIKVFSYSLVS 177
 145 GSGNLSADARSLQHNLYLFAFEENSNGNGCAIO-AQTELSRNVPISFARRADLNG 203
 178 GDGAIDQSLMVGNIEXLCTFEQENVADSGACQVTKTFSAVGKYPFLSLGVNAGKRG 237
 204 GAIC-----SNLDCGNVPLFPFGNSAT-- 228
 238 GGVAAVNDGQAGGATDLSVFANNTAVEFEGNSARIQGIYSDGNIS--FLGNKATVF 294
 229 -----NGAICISDLNTESEKSLIACQETLIFASNS 261
 295 LSNVAPSYVDPAAGGPPADKDNVGDGGAIFCNDTNI--GEVSFKDGVVFFSINI 351
 262 AKKGGATYAKHMYLRNGVPSFINNSAKIGALAIQSGSLIAGSGVLFQNNQRT 321
 352 AAGKGAIVAKKLTISDCGPVQFGLGNVANDGAIYLVQGGELISADRGDITFDGNLRM 411
 322 SDQG-----LVNAYILEKDAIISLEARNG-DILFDPIVOESSKESPLPSLOA 372
 412 ATQGAATYHDVWVNASNAISMTAGQITTLRAKKEGRILLFNPIEN----- 456
 373 SVTSPTPATASPLVIGTSANR-----SVTFSESLSEBEKTPDNLTSQLQOPIFLKSGR 426
 457 -----ANGQPVIGTLTVNEGEGETTGDIVPAK-----GDVY--LVYSIELSOGR 497
 427 LVKDRVLSAPSLSDPQALLIMEAGTSL-----KTSDDLKATLSIPLSHSDTEKS 479
 498 IILREOTKLVNLSLTQTGGS-VHMEGSLDQFAVTTTPAANSMLTNYHSLASLKNNG 556
 480 VTIHAPNLSIQ-----KIFLSNSGDEFENVELLSKEONNIPLLT 520
 557 VTNPPTNPVQVSSPAVIGNTAGTIVTISGPIFFEDL-DETFAYDNQNMVLGADQ-----T 609

RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-J138:
RX MEDLINE-20330349; PubMed-10871362:
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.:
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AE001585; AAD18163.1; -
DR EMBL: AP002545; BAA98215.1; -
SQ SEQUENCE 922 AA: 100457 MW: DFF2AB633AB031C CRC64:

Query Match 16.3% Score 740; DB 2; Length 922;
Best Local Similarity 27.2% Pred. No. 3 Le-44;
Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

OY 103 FFFLSFTDCSSKSSPSIIHOKNGQLSRNNGSMSCRNHAECSGGAISADAFSLQHNLY 162
DB 123 FSTLSIQSPGDIKEGCLYSKNA-LMLNNYVYREFQNSKKGALISGANVTIVGND 181
OY 163 FTAFEENSSKNGALQAOFTSLSRNVSPISEFARNRA-DINGAIC-C-SNLICSGVNP 220
DB 182 SVSFYONATFGCAIHSSGPIQIAVNOAEIRFQONTAKNGSGALVSDGIDIDQNAAYL 241
OY 221 FFTGNSAT-----NGAICISDNTSEKSL-SLACNOETLFASNAKEKGAIVAKHMV 275
DB 242 FHENEALTTAIGKGAVALCCLPTSGSTPVPYVTFSDNKQVLFERNHISIMGGAIVARKLS 301
OY 276 LRYNGPVSEFIN-----NSAKIGAIQIOGGSLSILAGEGVLFYQNNNSORTSDOGLVRNA 330
DB 302 ISSGCTFLINNIYVANSQNLGALAIIDTGBEISLSEKGTIFQGN--RTSLPFL--NG 357
OY 331 IYLEKDAIILSSLEARNG-DLFFDPPIVOESSKESPLPSSLOASVTSPPATASPLVIOT 389
DB 358 IHLQNAKFLQARNGSYIEFYDPITSEAD-----GSTQLINQDPK-----NK 402
OY 390 SANRSVIFSESERLSEEEKTPDNLTSOLOPIELKSGRLVKDRAVLSAPSLSDOPALLI 449
DB 403 EYTGTLIFSGEKLAND--PRODKSTIPQVNVNSAGLYVKEGAETVYSKFTSPGSHLV 460
OY 450 MEAGTSLKTS-SOLKATLISPLHSIDTEKSVTIHAPNLSIORKIFLSNGD-----ENFY 503
DB 461 LDGCTLLIASKEDIAITGALAIIDISLSSSTAIVIKANTANKOISVTDSELISPGNAY 520
OY 504 ENVELLSKEDONNIPLLTSLK-EQSHLHPDGN--LSHFYOGDWTFSKSDSDE--GHS 557
DB 521 EDLRM--RNSQTPPLSLLEPGAGSVTVAGDPLVPSPHGFGQNMKLAWTGTGNKVGEE 578
OY 558 LIANMPKNTVPPEPQSTLVANTLWNTYSDMQAVOSMITIAHGAYLFTGTGSAVSNL 617
DB 579 F---WQKINKPRPEKEGNLVPIILMGNAVDSLMQVOETHASSLOTDRGLIDGIGNF 635
OY 618 FYAHDSGKPIIDMHHSRLCYLFCISTHSLDHSPLAACQLCKSSDSPTSTETTSYI 677
DB 636 F--HYVASENINIRYRNSGCVLTVSNNEITPKHTYSMAFSQLFSRDKDYAVSNNEIRMYL 693
OY 678 ATVOAQLATSLMKISAQA-----CYNESIHELTKYRS 710
DB 694 GSYLYOYTTSLGNIIFRASNPNVNVGILSRREFLQNPMLIFHLCAVGAHTNMKTDYAN 753
OY 711 FSEYEGCSHMSVAVSGVCSIPIVNSGCG-LFSSPSIFSKLOGFSGTQCGFESSGSEIR 769
DB 754 FPAVK-NSMNNNCWAIECGSMPLVFEENGRLEFGALPEAKLQDLYVAYQDFETADGR 812
OY 770 SFASASSFRNLSLPIGITEFEKSKQTRTYVFLGAYIQDLKRDVESGPFVLLKNAVSADP 829
DB 813 RFNSGSLTISVPLGIRFERLALSQDLYDFSFYIPDIRKDPSCCAALVIGDSMLVP 872
OY 830 MANLDSRAVYFRLTNORALHRLQTL---NVSCVLRGQSHSYSLDGTYYRF 878

DB 873 AAHVSHAFVSGTGRYHENDYTELLCRGSIEC--RPHARNYNINGSKFRF 922

RESULT 24

ID O9KIY9 PRELIMINARY: PRT: 922 AA.
AC O9KIY9:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN CP0770.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Unayam L.A., Ulfenback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwyn M., Nelson W., Deboy R., Kolonay J.,
RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.:
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AE002237; AAF38570.1; -
DR TIGR: CP0770;
SQ SEQUENCE 922 AA: 100480 MW: 3442106888AB0FA6 CRC64:

Query Match 16.3% Score 739; DB 2; Length 922;
Best Local Similarity 27.2% Pred. No. 3 Le-44;
Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

OY 103 FFFLSFTDCSSKSSPSIIHOKNGQLSRNNGSMSCRNHAECSGGAISADAFSLQHNLY 162
DB 123 FSTLSIQSPGDIKEGCLYSKNA-LMLNNYVYREFQNSKKGALISGANVTIVGND 181
OY 163 FTAFEENSSKNGALQAOFTSLSRNVSPISEFARNRA-DINGAIC-C-SNLICSGVNP 220
DB 182 SVSFYONATFGCAIHSSGPIQIAVNOAEIRFQONTAKNGSGALVSDGIDIDQNAAYL 241
OY 221 FFTGNSAT-----NGAICISDNTSEKSL-SLACNOETLFASNAKEKGAIVAKHMV 275
DB 242 FHENEALTTAIGKGAVALCCLPTSGSTPVPYVTFSDNKQVLFERNHISIMGGAIVARKLS 301
OY 276 LRYNGPVSEFIN-----NSAKIGAIQIOGGSLSILAGEGVLFYQNNNSORTSDOGLVRNA 330
DB 302 ISSGCTFLINNIYVANSQNLGALAIIDTGBEISLSEKGTIFQGN--RTSLPFL--NG 357
OY 331 IYLEKDAIILSSLEARNG-DLFFDPPIVOESSKESPLPSSLOASVTSPPATASPLVIOT 389
DB 358 IHLQNAKFLQARNGSYIEFYDPITSEAD-----GSTQLINQDPK-----NK 402
OY 390 SANRSVIFSESERLSEEEKTPDNLTSOLOPIELKSGRLVKDRAVLSAPSLSDOPALLI 449
DB 358 IHLQNAKFLQARNGSYIEFYDPITSEAD-----GSTQLINQDPK-----NK 402
OY 390 SANRSVIFSESERLSEEEKTPDNLTSOLOPIELKSGRLVKDRAVLSAPSLSDOPALLI 449
DB 403 EYTGTLIFSGEKLAND--PRODKSTIPQVNVNSAGLYVKEGAETVYSKFTSPGSHLV 460
OY 450 MEAGTSLKTS-SOLKATLISPLHSIDTEKSVTIHAPNLSIORKIFLSNGD-----ENFY 503
DB 461 LDGCTLLIASKEDIAITGALAIIDISLSSSTAIVIKANTANKOISVTDSELISPGNAY 520
OY 504 ENVELLSKEDONNIPLLTSLK-EQSHLHPDGN--LSHFYOGDWTFSKSDSDE--GHS 557
DB 521 EDLRM--RNSQTPPLSLLEPGAGSVTVAGDPLVPSPHGFGQNMKLAWTGTGNKVGEE 578
OY 558 LIANMPKNTVPPEPQSTLVANTLWNTYSDMQAVOSMITIAHGAYLFTGTGSAVSNL 617
DB 579 F---WQKINKPRPEKEGNLVPIILMGNAVDSLMQVOETHASSLOTDRGLIDGIGNF 635

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:25:16 ; Search time 14.17 Seconds
(Without alignments)

2122.535 Million cell updates/sec

Title: US-09-677-752-4

Sequence: 1 MRPDHNMFCICAAIISSTA.....CULRGOSHSYSLDITQTVRF 878

Scoring table: BLOSUM62
Gapop 10.0 , Gapexl 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	475	10.5	964	1	OMPE.CHLTR
2	451	9.9	1034	1	OMPE.CHLTR
3	197.5	4.4	1250	1	YFAL_ECOLI
4	175	3.9	1140	1	YMG9_YEAST
5	167	3.7	1609	1	FIG2_YEAST
6	149.5	3.3	1306	1	MSB2_YEAST
7	148.5	3.3	995	1	Y109_YEAST
8	144	3.2	725	1	AGAL_YEAST
9	143	3.2	1169	1	YK82_YEAST
10	139.5	3.1	983	1	Y144_HUMAN
11	138	3.0	469	1	AMY_CLOAB
12	135	3.0	1802	1	HKRI_YEAST
13	134.5	3.0	1643	1	OMPB_RICPR
14	132.5	2.9	2319	1	FAB_MOUSE
15	131.5	2.9	918	1	YMBJ_CABEL
16	130.5	2.9	2845	1	APC_MOUSE
17	129	2.8	1128	1	BEM3_YEAST
18	128	2.8	959	1	N100_YEAST
19	127.5	2.8	956	1	YEF3_YEAST
20	127.5	2.8	1113	1	N116_YEAST
21	126.5	2.8	682	1	NISP_LACLA
22	126.5	2.8	881	1	YJH8_YEAST
23	126	2.8	1379	1	YFL9_SCHPO
24	126	2.8	1589	1	CC25_YEAST
25	125.5	2.8	775	1	BUD8_MOUSE
26	125	2.8	603	1	BUD8_YEAST
27	125	2.8	2156	1	ORP1_HUMAN
28	125	2.8	2842	1	APC_RAT
29	124.5	2.7	773	1	CDH_PHACH
30	124.5	2.7	2249	1	P190K_RICRI
31	123.5	2.7	547	1	FIBL_ADE40
32	123.5	2.7	1694	1	ICAO_HAETIN
33	123.5	2.7	1702	1	ICAZ_HAETIN

ALIGNMENTS

34	123	2.7	1477	1	KEIC_DROME	004652 drosophila
35	122.5	2.7	817	1	YG4A_YEAST	P46949 saccharomyc
36	122.5	2.7	1041	1	EGT2_YEAST	P42835 saccharomyc
37	122	2.7	1394	1	HAP_HAETIN	P45387 haemophilus
38	122	2.7	1656	1	OMPB_RICJA	006653 r outer mem
39	121.5	2.7	630	1	MUC1_MOUSE	002496 mus musculi
40	121.5	2.7	772	1	SYR3_UREPA	099433 ureaplasma
41	121.5	2.7	784	1	SP4_HUMAN	002446 homo sapien
42	121	2.7	562	1	FIBL_ADE41	P14267 human adeno
43	121	2.7	851	1	BCK2_YEAST	P33306 saccharomyc
44	121	2.7	1117	1	YN96_YEAST	P53753 saccharomyc
45	121	2.7	1569	1	YFUA_ECOLI	P52143 escherichia

RESULT 1
OMPE.CHLTR STANDARD: .PRT: 964 AA.
ID OMPE.CHLTR
AC 084877;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN E PRECURSOR.
GN PMPE OR CT869.
OS Chlamydia trachomatis.
OC Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.
OX NCBI_TaxID=613;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RX MEDLINE=9900809; Pubmed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusev R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.,
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001360; AAC68467.1;
CC
CC KW Outer membrane. Signal.
CC FT SIGNAL 19 964 POTENTIAL.
CC FT CHAIN 19 964 PUTATIVE OUTER MEMBRANE PROTEIN E.
CC SO SEQUENCE 964 AA: 104703 MW: 189986/D2B571CE2 CRC64;

Query Match 10.5%; Score 475; DB 1; Length 964;
Best Local Similarity 23.5%; Pred. No. 1.5e+22;

Matches 231; Conservative 143; Mismatches 342; Indels 268; Gaps 46;

QY	69	YVLGNSYCW-----FVSKLHITDP-KFALFKE--KGDLS-----IONFR-----	104
DB	8	FLIGNSLGLAREVPSRIFLMPNSVDPDKESLSKISLTGDTNHLJNLCYDLNLTALAI	67
QY	105	-----FLSTQCSKE-----SSPS-IHQKQO	127
DB	68	LQKTRNEGAANTYTDLSLFD-TQKEGIFAKNLTPESGAGCYASPSPTVEINDTGP	126
QY	128	LSLRNNGSMFCR-----NHAEG---SGGASADAFSLQHNLYLTFEEENSSKNGGAI	178
DB	127	VIFENN---TCRLFTWRNRYAADIKEGCAIHAQNLVYINHHDDVVGFMKKNFSYOGAI	183

DB 609 QILSIETTN----DQGLDSNMOSLNTLSLSPHYGYQGLMTPNNMTTYYTTTLNNSS 664
 QY 550 -----KSDSEG-----HSLIANMT 563
 DB 665 APTSATSIARQKKTSEFTTSNTTASIPNIKASAGSGSSASNSGEVITIKHTLVYNNMA 724
 QY 564 PKNVVPERPOSTLVANTLMNTYSDM-QAVOSMT--NTIAHGA--YLFCTWGSVAVSNL 617
 DB 725 PVGYIDPIRGDLIANSIVSHSGNNMTGRLPLPDNSWFLALQGAATLFTKQOKRLS-- 782
 QY 618 FYAHDSGKPTDKMHHMSLGLGISTHSDHSPCLAAQOLKSSDSDFITSETYSYI 677
 DB 783 YHGYSASK-----GYTVSSQASGAHGHKFLLSFSSQSSDKMKKEKTNNLSSRY 832
 QY 678 ATY---AOALATSLMKISAOCYNESTHETKTRSFSEKGFSGMHSVAVSG---EVCA 730
 DB 833 LSLALCFHPHEDRIALLGAAC-NYGHNMMSFYGT-KKSKGKFHSTTLGASLRCELRD 890
 QY 731 STPLVNSGGLFSSFSIFSKIOGFGTQDGFESSGGEIRSFSSASPRN-ISLPITGTFEK 789
 DB 891 SMPLRSLMLPFAQ-ALFSRTEPAS-----IRESGDLARLFTLQAHAVVSPIGIK-- 941
 QY 790 KSKOTRTYVFLGAYID---LKRDESG-----PVLLKNNVSMAPRANL 833
 DB 942 -----GAYSSDTWPTLSMEWELAYQPTLVYMKRPLNTLLIONGSMWTTNPL 989
 QY 834 DSRAYMFLTNQALHRLQTLNVCVLRGSHSVDLGD 873
 DB 990 AKHSFYGRGSHSLKFLKLFANVQAEVNTVSHYINAG 1029

RESULT 3
 YFAL_ECOLI STANDARD: PRT: 1250 AA.
 AC P45508: P45507; P45506; P39441; P76468; P77487;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HYPOTHEICAL 131.2 KDA PROTEIN IN UBIG-NRDA INTERGENIC REGION
 DE PRECURSOR.
 GN YFAL.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of Escherichia coli K-12".
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA Alpha H., Baba T., Fujita K., Hayashi K., Honjo A., Horiiuchi T.,
 RA Ikemoto K., Inada T., Isono K., Isono S., Itoh T., Kanai K., Kasai H.,
 RA Kasai H., Kim S., Kimura S., Kitagawa M., Kitagawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampaio G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
 RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP PRELIMINARY SEQUENCE OF 1-938 FROM N.A.
 RX MEDLINE=84272624; PubMed=6087316;
 RA Carlson J., Fuchs J.A., Messing J.;
 RT "Primary structure of the Escherichia coli ribonucleoside diphosphate
 RT reductase operon".
 RL Proc. Natl. Acad. Sci. U.S.A. 81:4294-4297(1984).
 RN [4]

RP SEQUENCE OF 925-1198 FROM N.A.
 RC STRAIN-K12 / EMG2;
 RA Estep P., O'Keefe T., Robison K., Church G.M.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 1180-1250 FROM N.A.
 RC STRAIN-OV6;
 RX MEDLINE=88201664; PubMed=2834621;
 RA Hussain K., Elliott E.J., Salmond G.P.C.;
 RT "The parD-mutant of Escherichia coli also carries a gyrA mutation.
 RT Mol. Microbiol. 1:259-273(1987).
 RN [6]
 RP IDENTIFICATION.
 RX MEDLINE=96032851; PubMed=7567469;
 RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,
 RA Danchin A.;
 RT "Detection of new genes in a bacterial genome using Markov models for
 RT three gene classes".
 RL Nucleic Acids Res. 23:3554-3562(1995).
 CC -1- SIMILARITY: TO E.COLI YDER.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS EXTENSIVELY FROM THAT SHOWN BY
 CC MANY FRAMESHIFTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AEO00313; AAC75293.1;
 DR EMBL: D90855; BAA16052.1; ALT_INIT.
 DR EMBL: D90854; BAA16050.1; ALT_INIT.
 DR EMBL: K02672; -; NOT_ANNOTATED_CDS.
 DR EMBL: U30459; AAA74094.1;
 DR EMBL: Y00544; -; NOT_ANNOTATED_CDS.
 DR Ecogene; EG12850; yfal.
 KW Hypothetical protein; Repeat; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 1250
 FT DOMAIN 919 948
 FT CONFICT 28 30
 FT CONFICT 40 40
 FT CONFICT 65 66
 FT CONFICT 431 431
 FT CONFICT 433 434
 FT CONFICT 478 478
 FT CONFICT 773 773
 FT CONFICT 853 853
 FT CONFICT 923 924
 FT CONFICT 948 994
 FT PAYQVLAQKGYGLNNIRANQAEMERRDAGGDCQTLN
 FT LRVIG -> LITSRC (IN REF. 4).
 SO SEQUENCE 1250 AA; 131152 MW; 17F98C05E299FC95 CRC64;

Query Match 4.4%; Score 197.5; DB 1; Length 1250;
 Best Local Similarity 19.7%; Pred. No. 0.0001;
 Matches 188; Conservative 143; Mismatches 372; Indels 249; Gaps 38;

QY 53 SLFALCAHASDDPLVYLGNSYCNFVSKLHTDKREALFKKGLUSTONFSLFTQCS 112
 DB 12 SLPSMINSILFANGVAAVTDSOCGYDVASCASROSLSGITOWMSINDGQWLVSDFMT 71
 QY 113 SKESPSTIHQKNGQLS---RNGSMSPCRNHAAG---SGGASADAFLSHNLTAF 166
 DB 72 NNASGAVFLQGAERFSLPENETGTLFANNTVGTGVNNGKAFKENSSTL-NLTDF 130
 QY 167 EENSSKNGCAIOA---OTFSLRNVSPISFARNR-DLNGCAI-CCSLIGSGNV--- 217
 DB 131 SCNVAGCYGALYSSTGNTDGTGAVDLRVTNAMFRNNIANGKCGALITYTINNVYLSDFD 190

RESULT 5
FIG2_YEAST STANDARD: PRT: 1609 AA.

AC P25653:
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE FACTOR INDUCED GENE 2
GN FIG2 OR YCR089M OR YCR89M OR YCR1102.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92397594; PubMed=1523889;
RA Wilson C., Grisanti P., Frontali L.;
RT "The complete sequence of a 6146 bp fragment of Saccharomyces
cerevisiae chromosome III contains two new open reading frames."
RL Yeast 8:569-575(1992).
CC - FUNCTION: REQUIRED FOR EFFICIENT MATING.
CC - INDUCTION: BY MATING PHEROMONES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL: X59720; CAA4254.1; -
DR PIR: S19504; S19504.
DR PIR: S25345; S25345.
DR SGD: S0000685; FIG2.
SQ SEQUENCE 1609 AA: 166049 MW: 7066AD7F85A7B852 CRC64;

Query Match 3.78; Score 167; DR 1; Length 1609;
Rest Local Similarity 19.48; Pred. No. 0.013;
Matches 177; Conservative 114; Mismatches 317; Indels 304; Gaps 35;

DB 84 ITDKKELFKKGLDIONFRFLSTDCSSKSSPSIHKNGQLSLNNGSMFCRHH 143
DB 235 LLSPLLEPLSSSGDLISLTQATINDOTSK-TIPTLVDAISSLPPLRSSM----- 286
DB 144 EGGCAISADAFSLQHN-----LTFAPENSSKNGCAIOAQ 181
DB 287 -----APTSGDSISHTSPPSKTSNGYDVLTNSIDPSLEFTTSEXSS-----TQ 333
DB 182 TFSLSR--NVPISARRNRADLNGAICCSNLICGNNPLFTFNSATNGAICISDL 239
DB 334 LSLLRASKSETVNFTASIASTPFGTDSATLI-----DPISSVSTAS--SPVGISRA 385
DB 240 NTSEKG-----SLSLACNOETLFASNSAKKKGKGA-----IYAKH 273
DB 386 NFGSGNSNYPESTASSSQYODMSSSLPLSOTTWVINTNQGASTTSPAYYST 445
DB 274 MYLRNGPVS-FI-----NNSAKIGALAIQSGSLIAGEGSVLEFONNSORTSDOG 325
DB 446 ATKYDGVITEVTCPLTQTKSAIGVSSISVPOASSFG-SLILSSNSTLAASN 504
DB 326 LVNRNIVLEKDAISSLERNDILFDPPIYGESSKESPL-----PSSLAQSVS 376
DB 505 VP-----ESTASGSSO-----YODMSSSLPLSOTTWVINTNQGASTS 545
DB 377 PTPATASPLVIOTSNRNVIFSSERLSEEEKTPDNLISLOOPIELKSGRLVLKRAVLS 436
DB 546 ----TTSRNVST-ATKY-----DGVITEVTCPLTQTKSAIGISS-----STIS 588
DB 437 APSLSODPOLLIMEAGTSLSKTSDDLKATLISLPLHSIDTEKSVYTH----- 483

DB 589 ATOTSK-PSSITILGIST-LQLSDATFKGTETINTHLMTSTSTETTFSGSDSYLC 646
DB 484 -----APNLIOKIFLNSGDENFENVELLSKEONNIPLLTSLK--EQSHLPLDGNL 535
DB 647 TSEVNLAASSLSVYPNFSSSGSTATTINSTVTGTSKYSTSVSNPTLASQ-----HV 700
DB 536 SSGEGYGDWTFSSKSDSEHSLIAN-----WPKNVYPHPEROSTLVANTLMNTSMDMA 591
DB 701 SSVNSLIDFT-----SNSTETIAVISNIHKTSNMDY-----SLTTQLKTSKQTL 748
DB 592 VOSMINTIANCGAYLFGTWSAVNSNLFYADHSSCKPLDMNHHRSLGYLFGISTSLDHS 651
DB 749 VLSTVTTTVCATVCAATVETTWCPA-SSIAVT-----TSISKTLVLTTEVCSHS 794
DB 652 FCL-----ACQLLCKSSDSFTTST----- 671
DB 795 ECTPTVITSVATSTTIPLSTSSSTVLSSTVSGAKNPASEVTINTQVSATSEATSTS 854
DB 672 -----ETTSTIATVQAOAT-----S 687
DB 855 TQVSATSATATASESTTSQVSTASETISTLGTQNTTGTSLPALSTEMINTTVSRK 914
DB 688 LMKISAQACVNESI-----HELKTFRSFSKEGCGSMHSAVSGEVC 729
DB 915 TLIISTEVCSSKCVPTVITEVNTSKTPNSGHSSTQLTEAVEVTLSSHOTVMTSEVC 974
DB 730 AS-----IPVNSGSLFSSFSIFSKLQCGSTODFEESSELSRFSASFRNLSLP 782
DB 975 SNSICTPVTIVTSVOMRSTFPFYLSTSSSLASTKSSLSLASEMSTFVSST---QSLP 1031
DB 783 IGTFEKKSQKT 794
DB 1032 LAFTCEKERTT 1043

RESULT 6
MSB2_YEAST STANDARD: PRT: 1306 AA.

AC P3234:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MSB2 PROTEIN (MULTICOPY SUPPRESSION OF A BUDDING DEFECT 2).
GN MSB2 OR YGR014W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;

RN
RP SEQUENCE FROM N.A.
RX MEDLINE=92383951; PubMed=1514328;
RA Bender A., Pringle J.R.;
RT "A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
RT defect."
RL Yeast 8:315-323(1992).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII."
RL Yeast 13:1077-1090(1997).
CC - PFM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC - SIMILARITY: SOME, TO YEAST HKR1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

OY 505 NVELLSKRONNIPULTLSKEQSHLHPDGNLSHFQYOD-----WTFSMKDSDECHSLI 559
DB 357 DL-LVRSKKEYP-----GTEFDISENDYTCNFNAEESAATSASI 397
OY 560 AMWTKNYPHBEROSTL-----VANTLMTYSMDQAVOSMINTIAHGAYLFGTWSGAVS 615
DB 398 YSFTPSFENSDISLTSKSKGGEVTFSPYSNGAF--SESNALIMG-----SVS 447
OY 616 NLFYAHDSGKPTDNMHRSLGLYFEGISTHSDHSCFLAAGOLLKSSDFTTSTETTS 675
DB 448 GLQRDDTEGS--VNN-----GEINLDNGSTYVVEFVSCKGTNIIISGMILYUH 494
OY 676 YIATVOAQLATSLMKISAQACYNESHELKTKRSPSKGFGSMHSAVSGEYCA----- 730
DB 493 YPDTFTGQ--TVVFKGEVL-----AVDPTETNATPIPVVGYTKKNOIATITALTSDG 548
OY 731 --SIPVSNQGLFS-----SFSIFSKLOG-FSGTODGFEESSGEIRSFSSASS 775
DB 549 TGGVLATGCGNRQFPAICGTFSSDPSVSEGFAGAYAYLYNNGVATSAASS 603

RESULT 8
AGAL_YEAST
ID P32323:
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-AGGLUTININ ATTACHMENT SUBUNIT PRECURSOR.
RN AGAL OR YNR044W OR N3431.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE-91304412; PubMed-2072914;
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT The AGAL product is involved in cell surface attachment of the
RL Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin."
RL Mol. Cell. Biol. 11:4196-4206(1991).
RN 12)
RP SEQUENCE FROM N.A.
RA Pohl T.M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -1- SUBUNIT CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL: M60590; AAA34382.1;
CC EMBL: Z71659; CAA96325.1;
CC PIR: S17031; S17031.
CC PIR: A41258; A41258.
CC SCD: S0005327; AGAL.
CC Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
CC Pheromone response.
FT SIGNAL 1 22 POTENTIAL.

```

```

FT CHAIN 23 725
FT DOMAIN 53 493
FT REPEAT 149 53
FT REPEAT 395 493
FT REPEAT 182 307
FT REPEAT 182 188
FT REPEAT 189 195
FT REPEAT 202 196
FT REPEAT 203 209
FT REPEAT 210 216
FT REPEAT 217 223
FT REPEAT 224 230
FT REPEAT 231 237
FT REPEAT 238 244
FT REPEAT 245 251
FT REPEAT 252 258
FT REPEAT 259 265
FT REPEAT 266 272
FT REPEAT 273 279
FT REPEAT 280 286
FT REPEAT 287 293
FT REPEAT 294 300
FT REPEAT 301 307
FT SEQUENCE 725 AA; 73353 MW; 70420C853B0B01F8 CRC64;

Query Match 3.2%; Score 144; DB 1; Length 725;
Best Local Similarity 19.6%; Pred. No. 0.11;
Matches 102; Conservative 85; Mismatches 223; Indels 110; Gaps 22;

OY 340 SSLRNAGDILFDPPIVQ--ESSKESPLPSLOAVSTPATASPLVIGTSANRSVIF 397
DB 156 SALEPSSASII--SPVSTLSSTSSNPTTSLSTSSPSTSSPSTSSPSTSS 213
OY 398 SSERLSEEXT--PDLTQLOLOPTELKSGRLVLRRAVLSAPLSODPQALLIMAGT 454
DB 214 SSSSTVSSSSSTSSPSTSSPSTSSPSTSSPSTSSPSTSSPSTSSPSTSS 273
OY 455 SLKTS--DLKLTATSLPLHSIDTEKSVTHAPNL-----STOKLFLNSGD--ENFY 503
DB 274 STSPSSKSTSAKSTSTSTSTSTSPSLTSSPPLASTSPSTSSPSTSSPSTSS 333
OY 504 ENVELLSKRONNIPULTLSKEQSHLHPDGNLSHFQYODWTFSMKDSDECHSLI 563
DB 334 SSSSTVSLYSPSTPVYSPVSSNVATPMSST-----VETVSSQSSSE-----YI 381
OY 564 PKNVPRPEROSTLVANTLMTYSMDQAVOSMINTIAHGAYLFGTWSGAVSNLFYAHDS 623
DB 382 TKSSI-----STIIPFSMTYF-----TVSGVTMTWTCP-----YSES 419
OY 624 SGRPIDMNMHRSLGLYFEGISTH--SLDDHSFCLAA--GOLLKS-----SDSFTSTE 672
DB 420 EFTLTLTSHET-----VTTDAITVCTHESCHPQOTTSLIISKMTKNVATSVSTV 472
OY 673 TTYIATVOAQLATSLMKISAQACYNESHELKTKRSPSKGFGSMHSAVSGEYCA 721
DB 473 ESSYACSTCAE--TSHSYSSVGTASSSSVTOQTSTKSWMSMTTSDDEPKHATGKYH- 529
OY 722 VAVSGEVCASIPVYNSGGLFSSSIFSKLOGFGTODGFEESSGEIRSFSAFNRISL 781
DB 530 -----VTSSTSTIS--TSVEASTSTSIDESQSHLSTLVLSSSSLA 575
OY 782 PIGITFEKKSKQTRTYFFLGAVIYODLKRDVESGPNVLK 821
DB 576 TL-----SSDST-----LLFSSVSSL--SVEQSPVTTLLQ 603

RESULT 9
YK82_YEAST
ID YK82_YEAST
AC P36170;
DT 01-JUN-1994 (Rel. 29, Created)

```

```

DB 300 SMLDPSQASLA-----QGLVPSNEMKOTAIQSPASGNTSESH-----SNVSMIGKC 345
OY 292 GCAIAIOSGGS-----LSTLAGEGSLFPNNQORTS--OQGLVRNMIYLE 334
DB 346 FCDVCEAKGSGTTGSOFLQFQTAQALQALAOHS---OSGTTTSSWMC----- 393
OY 335 KAILLSLEARNODLFPDPIVOESSKESPLPSLQASVTSPT--PATASPLVITQTSANK 393
DB 394 -----STIOS--PSLVQYOLKNSDSAVISPF-----TKR 421
OY 394 SVIFSSERLSE---EERTPDNLTSOL-----QOPIELKSGRLVLDKRAVLSAPSL---SQ 442
DB 422 QAFPTSTMEVEFLQEKSPAVATSTAPPPSSPLPSKS-----TSAPOMSPGSS 471
OY 443 DQOALLIMEAGTSLKSSDLKATLSLPLHSIDTEKSVTHAPNLISOKIFLSNGDENF 502
DB 472 DMOSSSPOPAQOKLKQOKKASLTSTKIPALAVMPCGADISGLNLOFGALQFCSEPVLSO 531
OY 503 YENVLLSKEONNIP--LLTLSKEQSHLHPDGNLSHFQYQ-----DWTFSWKDSEGC- 555
DB 532 YESTPTTSSSSQAPSSLTSTASSESSSTI--SSNOQESQYOGPIOSTTYTSQNNAGCP 590
OY 556 ---HSLIANWTPKYNVPPPEROSTLVANTLMWYSDMOAVOSMINTIAHGAYLFTGWG 611
DB 591 LVEQRSTOTRRPYSISSSPQKDLTOAK-----NGFSSVQATQLOTTQSVEGAT-----G 640
OY 612 SAVSNLEFYAHDSGKPID-----NMHRSIGYLFQISTISLD-----DHSF 652
DB 641 SAVKS--DSPTSSIPPLNETVSASALLTTNQHSSSLGGL---SHSEEPNPTTTTQHSS 695
OY 653 CLAAGO--LKGSSDSFTTETTSYIATVOAOLATSLMKISAQACNESIHELKTKYRSF 711
DB 696 TLSTQONTLSSSTISGRTISTL-----LHTSV-----ESANLHSSSTF 736
OY 712 SKGFGSMHSAVAGEVCASIPYNSGGLFSSFSIFSKLQGFSGTODGEESGGEIRSF 771
DB 737 S-----TTSSTVSAPPVSVSSSINSGLSLGLSL--GSNSTVTASTRSSVATTSG 785
OY 772 SASSFRNISLPIGI 785
DB 786 KAPP---NLPGV 795

RESULT 11
AMT_CLOAB STANDARD: PRT: 469 AA.
ID AMY_CLOAB
AC P23671;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PUTATIVE ALPHA-AMYLASE (EC 3.2.1.1) (1,4-ALPHA-D-GLUCAN
DE GLUCANOHYDROLASE) (FRAGMENT).
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 792;
RX MEDLINE=91072241; PubMed=2254264;
RA Getischer U., Duerre P.;
RA "Cloning, sequencing, and molecular analysis of the acetate
RA decarboxylase gene region from Clostridium acetobutylicum.";
RL J. Bacteriol. 172:6907-6918(1990).
CC -1- CATALYTIC ACTIVITY: ENDOMYDOLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M55392; AAA63759.1;
DR PIR: B37837; B37837.
KW Hydroxylase; Glycosidase; Carbohydrate metabolism.
FT NON_TER
SQ SEQUENCE 469 AA; 50870 MW; DB1B6702D9B01572 CRC64;

Query Match 3.0%; Score 138; DB 1; Length 469;
Best Local Similarity 19.0%; Pred. No. 0.14;
Matches 142; Conservative 81; Mismatches 214; Indels 312; Gaps 36;

OY 124 KNGQLSRNNGSMSCFRNHAEGSGAISADAFSLQHNYLFTAFEESSKNGGAIQAOQTF 183
DB 4 KSGDLH-----NAQGMGGLDSNKKCYDMLETH--DEYHNESK----- 38
OY 184 SLSRNVPISFARNRADLN-----GGAICCSNLICSGVNPPLF---TGNATNGAIC 234
DB 39 -----DLTDWQKAGMALIASR---AGSV--PLFFDRPTGNIGSEGDALM 78
OY 235 CISDLNTSEKGSLSLACNOETTFASNSAKKGAIIAKHVLVLRNGPVSTINSAKIGCA 294
DB 79 KDSDVAVVNEFHNAAGONEYLRLON-----NNKAMI--- 110
OY 295 IAIOSGGSILAGEGSLVFQNNQORTSDQGLVRNMIYLEKDAILSLEARNGLIEFDP 354
DB 111 -IERGSKGAVIVNEGDFNLTP-----INLEGND----- 141
OY 355 IVOESSKESPLPSLOASVTSPTPATASPLVITQTSANRSVIFSSERLSEEXTPDNLTS 414
DB 142 -NHGSAITSLYS--QGRNTGTVPANSIIVLYNNKNSNP----- 177
OY 415 QLOOPIELKSGRLVLDRAVLSAPLSODQALLIMEAG--TSKLTSSDLKATLSLPLHS 473
DB 178 -----SDRVTLSQOAKKAGDSVT-----ITYDAGTALKADASNVL-----YWG 216
OY 474 LDTEKSVTHAPNLISLQIKIFLSNGDENFENVELLSKEQNNIPPLTLTSKEQSHLHPDG 533
DB 217 YDGFSAATSKA-----MTSLGDKKKQOTTIVPEVTKNV-----NFSITDG 257
OY 534 NLSHFEGYQGDWTFSMKDSDEGHSILANWT---PKNVPPH---PEROSTLVANTLMNT 585
DB 258 T-----SW--DNNGG---ANWNIPLASNYLPHGGYKVYDDSSNLVSGNNFTI 299
OY 586 YSDMOAVOSMINTIAHGCAVLFTGWG-----SAVSNLEFYAHDSGKPIDNMHRSGLYLF 641
DB 300 YVNGNLANSS--NVSLLH-----WGVNGWSNMOLAMVYDSNC-----FW 336
OY 642 ISTHSLDDHS---FCLAAGOLGKSSDSFTTETTSYIATVOAOLATSLMKISAQAC 697
DB 337 EATTAIPASSFTLNFCTTNGSSMDNNNNNNMT--LNTSSVYPKVQ----- 379
OY 698 NESIHELKTKYRSFSGKFGSMHSAVAGEVCASIPYNSGS--GLFSFSIFSKLOGFS 755
DB 380 -----VTAPAECOKISVYVNGSLASASINITLHWGCGCT 415
OY 756 GTODGFEESGSEIRSFSASSFRNISLPIGITEFKKSKQRTYVYFLGAYIIDLKRDYEG 815
DB 416 SPOD-----INMVKQADGRMLANITLPSGC-----YVNVMAFMD-----QSG 452
OY 816 PVLILKANVSDAMANDSRAVYMERLTN 844
DB 453 -----TWD-----NNNSNNVNSSSTN 468

RESULT 12
HKRL_YEAST STANDARD: PRT: 1802 AA.
ID HKRL_YEAST
AC P41809;

```

RT "Characterization of the gene encoding the protective paracrystalline-
 RT surface-layer protein of Rickettsia prowazekii: presence of a
 RT truncated identical homolog in Rickettsia typhi.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BREINL;
 RA Moron C.G., Yu X.J., Walker D.H.;
 RT "Sequence analysis of ompB of Rickettsia prowazekii.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-MADRID E;
 RX MEDLINE-99039499; PubMed-9823893;
 RA Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 RA Scharitz-Ponten T., Almarik U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RL Nature 396:133-140(1998).
 RN (4)
 RP PARTIAL SEQUENCE.
 RC STRAIN-BREINL;
 RX MEDLINE-92114896; PubMed-1370573;
 RA Ching W.M., Carl M., Dasch G.A.;
 RT "Mapping of monoclonal antibody binding sites on CNBr fragments of
 RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
 RT prowazekii.";
 RL Mol. Immunol. 29:95-105(1992).
 RN (5)
 RP IDENTIFICATION OF CLEAVAGE SITE.
 RX MEDLINE-92104668; PubMed-1729180;
 RA Heckstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;
 RT Evidence for proteolytic cleavage of the 120-kilodalton outer
 RT membrane protein of rickettsiae: Identification of an avirulent mutant
 RT deficient in processing.";
 RL Infect. Immun. 60:159-165(1992).
 RN (6)
 RP FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
 CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
 CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
 CC - FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
 CC - CELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
 CC LAYER WITH HEXAGONAL SYMMETRY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M37647; AAA26390.1; ALT_INIT.
 DR EMBL: AF161079; AAD42234.1;
 DR EMBL: AF235273; CAA15140.1;
 KW Antigen; S-layer; Cell wall.
 FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
 FT 1329 1643 32 KDA BETA PEPTIDE.
 FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
 FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
 FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
 FT CONFLICT 178 179 AA -> VC (IN REF. 1).
 FT CONFLICT 191 201 TTOTAPLITGA -> INSRSSHYLVS (IN REF. 1).
 FT CONFLICT 212 212 T -> I (IN REF. 1).
 FT CONFLICT 313 313 Q -> L (IN REF. 1).
 FT CONFLICT 1104 1104 D -> G (IN REF. 2).
 FT CONFLICT 1123 1123 T -> S (IN REF. 2).
 SQ SEQUENCE 1643 AA; 169854 MW; 735FDF392E6346CC CRC64;

QY 130 LNNNGS---MSECRNHAEGSGAISADEFSLQHNLYLFTAFEENSSKNGCAIOAOTFSL 185
 Db 424 VKNNGNAGVITFNAN-----GALVSAST-----DENIVATNIAIEAGAGV 466
 QY 186 SRNVSPISFARNRADLNGAI---CCSNLCSGVNPLFFTGNGATNGAICCSIDNTSE 243
 Db 467 V-ELSGIHIAELRLG-NGGSIFFKLADGTVINGPNNALNNALAAAGSI---QLDGSA 520
 QY 244 -----KGSLSLACQNETLFASSAKKEGCAIYAKHMYLRYPGVSEFINNSAKIGGAIA 296
 Db 521 IITGDINGGVNALOHLITL-ANDASK-----ILALDG-ANLT--GANVGAIH 565
 QY 297 IO-SGGSLSILAGEGVLPFONNSORTSDGLVRNATYLEKDAILSLSEARNGOILFFPDPT 355
 Db 566 FQANGGTIKLNTQNNLVNFFDLDTITDKTGVDASSLTNNQTLTI-----NGSI----- 615
 QY 356 VOESSKESPLPSLOASVSPPTATAPLYIOTSANRSVIFSSERISEEKPDLTSQ 415
 Db 616 -----GTVAANTKTTLAQ---LNIGSSKTLTNAGDVAINELVIEENGSVQ 656
 QY 416 LQPIEL--KSGRLVKDRAVLSAPSLSDPQALLINEAGTSLKTS---SDKLATLSI 469
 Db 657 LNHNTYLIITKTINANOGQILVADPLNTN---TTLADGTNLGSAENPLSTIFATKAA 712
 QY 470 PLS-LDTEKSVTIHAVNLISLOKIFLSNGDENFENVVELSKBONNIPULTLSKEQSH- 527
 Db 713 NADSLVNGKVNLYANNITND---ANVGLHRSG-----GTSIVSGTVGGQGHK 762
 QY 528 ---LHLPDGLSSHFYOGDWTFSWKSDSGSL-IANTPKRYVHPPEOSTL-VANT 581
 Db 763 LNNLILNGTTVK---FLGDTTFNGGTRITECKSLIOISNNYTTVHVESADNTGTLFVNT 819
 QY 582 LMTNYSMDQAVOSMINTIAHGAGLFGWGSV---SNLFYAHDSGKPIDNMHRHSLG 637
 Db 820 -----DPTVTILNKGAY-FGLYKOVYIISGONIFY-----NELGN-----VG 856
 QY 638 YLFGTSHSLDHSFCIALACGLLKKSSDSPTSTETTSYATVQAOALATSLMKISAQCY 697
 Db 857 IVHGIANNSIFENASLGLSLFLPSGTLVDLTIKSTGVNTQVNFNPIYVVGSDSMI 916
 QY 698 NESIHELTKTRFSKSGFGSGWHSVAASGEVCASPIYNSGSLFSSSISISKQGPSGT 757
 Db 917 NNG--QILGDKKNIIIALSGDSNITV-----NANTLYSGIRTKNNGVTVL 962
 QY 758 QDGFESSGELRSFSASSFRNLSLPIGITFEKKSOCTR---TYYPFLGAYIO----- 806
 Db 963 SGMKPNFQTL-----YGLGLENGSKKLKOVFTTDYNNLGSIIANNVTIN 1008
 QY 807 -----DLK-----RDVES-----GPVLLKNA 823
 Db 1009 DVTITTTGCIAGTDFDAITILGSVNGNANVFEDSTFSDPSMLIATQANKGTVYLLNA 1068
 QY 824 V-----SMDAPMANLDSRAVFRLTNORALHRLQTLNLVSCYLKQOSYSYSLDGT 874
 Db 1069 LVSNIGSLDTPVASV-----RFTGN-----DSGALQGNIVYSONIDFQT 1107
 RESULT 14
 FAS-MOUSE STANDARD: PRT: 2319 AA.
 AC 006194;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COAGULATION FACTOR VIII PRECURSOR (PROCOAGULANT COMPONENT).
 GN C8 OR F8C.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.

DB 1343 TCCITKOID--HSKMKRF--ITKSSLDSSVYKSTTQT--NSSDSHIYKTSAFPIDLK 1396
 OY 681 OAOQLATSLMKISAOACYNESIHLEKTK 707
 DB 1397 RSPFQNKFSHYOA---SSYIYDFKTK 1419

RESULT 15
 YMOB_CAEEL STANDARD: PRT: 918 AA.
 ID YMOB_CAEEL
 AC P34487;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE HYPOTHETICAL 96.6 KDA PROTEIN F59B2.12 IN CHROMOSOME III.
 GN F59B2.12.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE-94150718; PubMed-7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 Fullon L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 Johnston L., Jones M., Kersey P., Kistner J., Laister N.,
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifkin L., Roopla A., Saunders D., Showkeen R.,
 Sims M., Smalton C., Smith A., Smith M., Sonhammer E., Staden R.,
 Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 Watson R., Watson A., Weinstock L., Wilkinson-Sprat J.,
 Wohlschlag P.;
 RA 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RA elegans.
 RT Nature 368:32-38(1994).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL/211505; CAAT7581.1;
 DR PIR: S31132; S31132.
 DR Wormpep: F59B2.12; CE01024.
 DR Hypothetical protein.
 KW SEQUENCE 918 AA; 96560 MW; E464FD86B14945DE CRC64;
 SQ

Query Match 2.9%; Score 131.5; DB 1; Length 918;
 Best Local Similarity 19.7%; Pred. No. 1;
 Matches 154; Conservative 114; Mismatches 311; Indels 201; Gaps 35;

OY 93 KEGDLSIQNRFSLFTDCSSKES-----SPSIIHQKNGQLSLRNNGSMFCRNHAECS 146
 DB 27 KAKFSSSSSGGASMFSGAKSSFGGFAMPKLDASKAAAHV-----SSSHKHGHSS 81
 OY 147 GCA-----ISADAFLOAH-----YLTFAPFENSSKNGCAI----- 178
 DB 82 GSSSNTHSLTYVAGDGKNTENSEKKDYNNESKVDENENTKIKSADGSVLETGKSHNK 141
 OY 179 ---OQOTFSLRNVPISFARNRDLNGAICCSNLICSGVNPPLFTGNSATNGAIC 235
 DB 142 SSDASSTGLEK-----SKTYADKNC-----TMLSSTNKKI-----NNGSRSAAL-- 182
 OY 236 ISDLMTSEKGLSLACNOETIFASNSAKKGAIIYAKHNVLYRNGPV-----FLN 286
 DB 183 -----DEGNEFVNQOANDGTFLRNNTGKNTDEHLSHVLDENAMQMSIGADGTSHTNITN 236

OY 287 NSAKTIGA--IAIQQSGLSLILAGECVLFPNNQORTSOGLVNRNATYLEKDAIISLEA 344
 DB 237 RKSGVDSHNAASDAHSNFEISLDAQCKKSKQNSKAAAS--GSMADF---ESNLESLKN 292
 OY 345 RNCDDL-----FFDPIVDESSKESPEPLSSLOASTSPPTASPLVYIOTSANRSVFS 398
 DB 293 ADGTSMSNSTGNFNNTSYKATAEYMSKKNNVADGTSMEAS-----HAGSNSSKINS 346
 OY 399 SERLSEEEKT--PDNLTSLQQLFELKSGRLYLKDBAVLSAPLSQ-----DPOAL-- 449
 DB 347 ASGQSSDLMSVGPNGIKS---HSTSKTDNYAL--DEANOSAGSISQIDCKNGORSLNESS 402
 OY 450 MEGATSLKTSDDLKATLSIPLHSLDTEKSV--TIAPNUSIQKIFLNSGDEFYENVEL 508
 DB 403 IESGKRAESRNN---TAADTLDSVANGTVSSSHSKSAS-----GTSIDEN----- 445
 OY 509 LSKQNNIPLLTISKQSHLHPDGNLSH-----FGYGDWTFESMKDSDEG 555
 DB 446 ---HNKTHALQASVDE-----HGMKKNHSIDGSYRNKKTGEFG--NSEKASIKNADGT 494
 OY 556 HSLIANTPKNYVPHPEROSTLVANTLMN--TYSD-MQAVQSMINTIAHG-----GAYL 606
 DB 495 MSQVNYKNDTNRNTYEAKSALEKNEKNSDGTFFKDESKSNSRVNRTOGSGNLAAGSYS 554
 OY 607 FGTWG-----SAVSNLFVAHDSGKPIDMWHRSGLCYLFGISTHSLDHSFCLAGQL 660
 DB 555 VKGQGVSNFTIASSNAFNNTSDAESNOFDLHQKT-----ANCTEI 595
 OY 661 GKSSDSTTSTETTTATYATQAOALATSLMKISAOACNESIHLEKTKYRFSFEGESWH 720
 DB 596 THAKDS-----KOVAASANKKSSLDTSMSAVDAKG-----NKVDKTSQQAADSHD 640
 OY 721 SVANSGVCASIPIVNSGGLFSSFSIFSKLQSGTQODFEESGCEIRFSFASFNIS 780
 DB 641 AIASSPVDKAKYKHADRS---ESISNDSNQTAASEHNDSSKQSEHEKRONADGSFSDVS 697

RESULT 16
 APC_MOUSE
 ID APC_MOUSE STANDARD: PRT: 2845 AA.
 AC Q61315; Q62044;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE APOEMNATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN) (MAPC).
 GN APC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1 AND 2), AND VARIANTS.
 RC STRAIN=C57BL/6J; AND CAST/EI; TISSUE=Brain;
 RX MEDLINE=92263101; PubMed=1350108;
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 Luongo C., Gould K.A., Dove W.F.;
 RT "Multiple intestinal neoplasia caused by a mutation in the murine
 RT homolog of the APC gene";
 RT Science 256:668-670(1992).
 RL
 RN [2]
 RP ERRATUM.
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 Luongo C., Gould K.A., Dove W.F.;
 RL Science 256:1114-1114(1992).
 RN [3]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Liver;
 RA Dicker F., Lambertz S., Reilmair A., Balhausen W.G.;
 RT "The murine APC gene: alternative splicing of 5' untranslated
 RT region segments";
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 RN [4]

Query Match 2.8%; Score 126; DB 1; Length 1589;
 Best Local Similarity 18.1%; Pred. No. 5.2;
 Matches 162; Conservative 141; Mismatches 322; Indels 272; Gaps 37;

112 SSKSSPSTIHKKNOLSRNNGSMFCNNAHGGGGAISADA-PSLOHNYLFTAEENS 170
 143 SSKSSRRSLN-----SLGNSAYLHVPRNPSKSRGSSSTLASUSNAIN-----AETSS 191
 171 SGNCGAIAQAFSLSRNVPISFARNRADLNGAICCNLICSGVNPDLFTGNSATNG 230
 192 GHNNTYMANNSPFSAPNDASHIT--PQSNFNSMALSODMKRSAD-----GSEEMNT 242
 231 GAICCSIDLNTSEKSGSLSLACNOETLFASNSAKKGAIAAHMV-----LRYNCP 281
 243 NA1-----MNNNE-----TNLQT-----SGKAGPPLVAETIKILPLEETEMIING- 284
 282 VSFNNNAKIGCAIAIQSGSLSILAGEGVLFONNSORTSDOGLVNMAYLEKD----- 336
 285 --IRSNIASTWSPILPLIT-----RTSDRYLV-----YNNKDLTYC 318
 337 ---AIISSLEARNGDILFTDP-----IVQESSKESPLSSLOASVTPPTATAS 383
 319 SELPLISNIMESDDICDEPKPPPDHLYNLYTRDLRKNANIEDSSTSKOSEQONS 378
 384 PLYIQTSAHSRVIFSSERLSEEEKTPDNLTSOLOQ--PIELKSGRLVLDRAVLASPSLS 441
 379 SILMEKQSGKETDGGNNNSINDOONNNKNEFENAGPSSLNS-----LSAPDLT 428
 442 QDPQALLMEAGTSLKTS-----DLKATLSTPLSLDTEKSVTHAPLWLOKITF 433
 429 QNIOSRVVAPRSISILAKSDIFHYHSRDIKLTW--ELDOLTYVYKTAH-----KMF 478
 494 LMSNGDENFENVELLSKE-----QNNIPLTLTSKE-----Q 525
 479 LKEN-RLNFYKYDILSDSVFTFOLGCRMLQHEIKAKSCKEIKKIFKGLISSLSRISIN 537
 526 SHLPLDGNLSHFYQGDWTFMSKDS-----EG-----HSLIA----- 560
 538 SHLY-----FDSA FHRKKMDTMNDKNDQENNCSPTEGDDCKIEVDVHDLVSPLSGK 592
 561 -----NMV-----PKNTVPRPEKOSTIVANTLNTYS--- 587
 593 RNVTSTDTLTPMRSSFTVNENDEMNFVLPGRNSVNSVWTPRISTONSTLEDFSPSN 652
 588 -----DMQAVOSMINTIAHGCAYLFGTWGSAVSNLEFVHADSSGKPRIDMHHRSGLYLF--G 641
 653 KNFKSAKSTIEMDV-----EESKFLRHVOLLIFVLQSSVSDDTLRQLLPREFKGS 705
 642 ISTHSLDHSFLCAAGOLLKSSDSFTTET-----SYIATVQALATSLMKISAQA 695
 706 FSGSGMWNPFSTFTDEFGNATKKAATSNBVAASSKNSISIRPKMADALIASAGYS 765
 696 CYNESHLELTK-----YNSFKEGFGSMHSAVSGEVCASITVSGSLF--SS 744
 766 ANSETNSQIDLKASSAASGVFTPFNRPSSHNRFTSRAVSRKRRKRYLTVDTLNTMKKS 825
 745 FSTPSKLOGFCGQDDFEESGGEIRSPASASFRNISLPDITGFEKKSKOKRTYVYFLGAY 804
 826 SOIEKLNMTG-----EHLKITSKPKSRIRHLEIN-SSTYQINO-----NVL 868
 805 IODLRDVEGPPVLLKNAVSMAPMANLDSRAYMFRLLNORALHRLQTLNNSCVL 861
 869 LLELLENLDSLIFNLKNLITPSILLDLSEEF-----VHAMSVSVSLV 914

DE HOMEOBOX PROTEIN SIX4 (SINE OCULIS HOMEOBOX HOMOLOG 4) (SKELETAL
 DE MUSCLE-SPECIFIC ARE BINDING PROTEIN AREC3).
 GN SIX4 OR AREC3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=Skeletal muscle, and Myoblasts;
 RX MEDLINE=96178400; PubMed=8628654;
 RA Kawakami K., Ohto H., Ikeda K., Roeder R.G.;
 RT "Structure, function and expression of a murine homeobox protein
 AREC3, a homologue of Drosophila sine oculis gene product, and
 RT implication in development."
 RL Nucleic Acids Res. 24:3103-3110(1996).
 CC -I- FUNCTION: INVOLVED IN SKELETAL MUSCLE DEVELOPMENT. ALSO IMPLICATED
 CC -I- IN RETINA AND KIDNEY DEVELOPMENT.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
 CC -I- ALTERNATIVE PRODUCTS: AT LEAST THREE ISOFORMS OF THIS PROTEIN ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. M8 CONTAINS NO HOMEOBOX DOMAIN
 CC AND IS NOT FOUND IN MUSCLE.
 CC -I- TISSUE SPECIFICITY: EXPRESSED IN DEVELOPING SKELETAL MUSCLE. ALSO
 CC PRESENT IN ADULT RETINA AND IN THE DISTAL TUBE OF KIDNEY.
 CC -I- SIMILARITY: BELONGS TO THE SIX/SINE OCULIS FAMILY OF HOMEOBOX
 CC PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D50416; BAA08915.1; -
 DR EMBL: D50417; BAA08916.1; -
 DR EMBL: D50418; BAA08917.1; -
 DR MGD: MGI:106034; S1x4.
 DR InterPro: IPR001356; -
 DR Pfam: PF00046; homeobox; 1.
 DR PROSITE: PS00027; HOMEOBOX_1; 1.
 DR PROSITE: PS50071; HOMEOBOX_2; 1.
 KW Developmental protein; Homeobox; DNA-binding; Nuclear protein;
 KW Alternative splicing.
 FT DOMAIN 2 5
 FT DOMAIN 58 61 POLY-SER.
 FT DOMAIN 70 76 POLY-ALA.
 FT DOMAIN 92 95 POLY-ALA.
 FT DNA_BIND 216 275 HOMEOBOX.
 FT DOMAIN 582 775
 FT VARSPPLIC <1 9
 FT VARSPPLIC 37 100
 FT VARSPPLIC 188 319
 FT VARSPPLIC 320 775
 FT SEQUENCE 775 AA; 82262 MW; B06EBB64E04E5061 CRC64;
 SO

Query Match 2.8%; Score 125.5; DB 1; Length 775;
 Best Local Similarity 20.5%; Pred. No. 1.9;
 Matches 147; Conservative 67; Mismatches 219; Indels 283; Gaps 37;
 3 PDHNFCCICAAI-----LSSTAVLFGODPLGETALLTKNHNVCYTFE 47
 103 PDHV-ACVCEALQGGNLDRLARFLWLSLPSDDLRCGNSILLKARALVA-----FH 151

```
OY 48 DCTMESLFALCAHA--SODDPLVVLGNSYCMFVSKLHTDPKEA----- 90
Db 152 OGIPPELYSILESHSFESANHPLL---QOLWY--KARYTEARARGPLGAVDKYRLRR 205
OY 91 -----LJKEKG-----DLSTIONPFF-----LSFTDCSS-- 113
Db 206 KFPUPRTIMDEEETVYCKEKSRNALKELYKON-RYPSPAEKRLAKITGSLTOVSMMF 264
OY 114 -----KESSPSIIHOKNGQLSRNNGSMSFCRNHAEGSGAISADAFSLQHNLYLTFAPEE 168
Db 265 KNRRQORBNPSETOSK-----SESDGNPSTEDESSKGH-----ED 299
OY 169 NSSKNGGAIQAOT-FSLSRNVSP1---SFARNRADLNGAICCSNLICSGNVNPLEFTTG 224
Db 300 LSPHPLSGSGSDGVNLSLSHVEPYMOQIGNAKISLSSGVLLNGSLVPASTSPVFLNG 359
OY 225 NSATNGCAICISDLN-----TSEKSLSLACNOETLEAS 259
Db 360 NSFIOGHNGVILNGLNGNTQTIVSLNPPKMSNIVNGIAMTDILGSTSDVKEKVLQS 419
OY 260 NSAKEKGCAIYAKHMYLVRYNGPVSPFN---NSAKICG--AIAIOSGSLSLAGEGCVL 313
Db 420 SAVNSAATTSP-----SAPVSPGLIPCTEVKREGIOQTVASQDGSV-----VT 465
OY 314 FQNNQSQRTSDQGLVR-----NAIYLEKDAITLSSLEARNGDLF---FDP1-VOESSSK 362
Db 466 FTTPVO-1NOYGIVQIPNSGANGQFL-----NGSIGFSPLOLPPVSVASOGN 512
OY 363 ESPLSSIAQAVTSPTPAT-----SPL-----VIOTSANRSVIF 397
Db 513 LSVTPSTSDGSTFTSEPTVHGKLFSLTPSAVVYTVPNSGOTVCAVKBQGLERGLVF 572
OY 398 -----SSERLSEEEKTPDNLTSQ-----LQOPTELKSG 425
Db 573 SOLMPVNHSAQVNASLSSENLSGGLHP--LTSILVNVSAHGFSLTPPTLLNPTELMPD 630
OY 426 RLVLKDRAVLAP-----SLSDPOALLIMEAGTSLKTSDDLKATLSIPL----- 471
Db 631 ---LAESQVPASAPVASKCTVSSVSNTNYATL--QNCSLIPGODL---LSGPMTQAALG 680
OY 472 -----HSLDTEKSVTIHAPNLSIOKIFLSNSGD--ENFYEVELLSKEONNIPLL 519
Db 681 EIVPTAEQVSHASTAVHODFVREORLVLOSVPNIKENFLONSF--NKATNNLMML 734
```

Search completed: May 6, 2001, 19:28:01
Job time: 165 sec

```

OY 291 -IGAIATOSGSL-----SILAGECVLPONNSORTSDOGLVRNAI-----Y 332
DB 402 ALESOSLNKSGELLEPSISDPMLEKMTDACKQTVESLSLKSDDATIPKPSVSLGVK 461
OY 333 LEKDAIILSLERN-----GDILFPDPIVOE----- 358
DB 462 LKNDASKCAVESOSLINDPSVSTSFYIPNMEKLNKAIEHQALNKAIALINKTOLP 521
OY 359 ---SSKESPLPSLOASVTSPTATA----- 382
DB 522 YFHOSSELPISAKRAALSOOTESASKSSNISSEKCDHPSPNSFISASQOAKDFESE 581
OY 383 ---SPLVITOSANRSVIFSSERISEE-----KPPDN-----LTSOL 416
DB 562 ATSSIDLILSPGRSLASLKLSTVKEAGARYNRSVATPVPMTSPNFEISKASQSLASTA 641
OY 417 COPIE--LKSGRVLKDRVLASPSISODPOLIMEAGTSLKTS-----SDLK 463
DB 642 OVPVEPAPKTSVNLKSLSLASKLS-----FOEAGTSLKERTANYQHSPEVLNDRS 694
OY 464 LATTISPLH---SLDTEKSVTHAP-----NLSTOKIFLSNGDENFYE- 504
DB 695 LASTLSARHADAISSSKOHTELAQPAFNTGTGLMTKTKNSLSDSALTPQHSDSVLPDI 754
OY 505 -----NVELSKRONNIPILITSKQSH-----LHLPDGNLSH- 538
DB 755 TSKARLASTCAKTDLHRPRKHKSFLENYFGHNAEDESPQSDVESQTPPSGYSEND 814
OY 539 -----FGYOGDMTFSWKD-----SDECHSLIANWTP-KNVVPHPEROSTLVANT 581
DB 815 ICGTDIYFPALOGAALAAKDFSRREASDSQASASNMMLPMKKLSASSSTRSTS 874
OY 582 LMNTYSDMOAVOSMIN-----TIANGA-----YLEGT----- 609
DB 875 LL-TYANQNHHSINETDIPMAAHIVATQONKKEKAPSPNDQALDHLGSAKRIP 933
OY 610 -----WGSVSNLFFAHDSSGKPIDNMHHRSL-GYLFGIS-----THLDDHSEFLANG 657
DB 934 SKKTESS--SSL---HIPSHNSFSDIHHRKIAHAHAALTAASEKLTPIESTYSSSKA 988
OY 658 QULGKSSDSFISTETTYATVQAOLATSLMKISAOACYNESIHLELTK-----YR 709
DB 989 HNASLSAALVANKKXLIHAPTPQV-LAPVLEKVSILPHSHSLSLOIREESEVEGVFD 1047
OY 710 SFS-KEGFGSHSV-----AVSGEVCASIPVNSGSLGSEFSIFSKLGFS 755
DB 1048 APSDKEDILGSSNAPLDVPHGANSRSMDEVNCGSKYDSDGILDGSGQYTD-----SDMS 1100
OY 756 GTQDGEESGGEIRSFSA 773
DB 1101 DDENSLKDSOSSVLSFSA 1118

```

```

RT Cyclase pathway".
RL Cell 46:789-796(1987).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE-86220116; PubMed-3011405;
RA Camonis J.H., Kalekine M., Condre B., Garreau H., Boy-Marcotte E.,
RA Jaquet M.;
RT "Characterization, cloning and sequence analysis of the CDC25 gene
RL which controls the cyclic AMP level of Saccharomyces cerevisiae.";
RL EMO J. 5:375-380(1986).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hallier L., Jier T.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevisan E., Vignati D.,
RA Wilcox L., Wohlman P., Yaudin M., Wilson R., Waterston R.;
RL Submitted (Nov-1994) to the EMBL/Genbank/DBJ databases.
RN (4)
RP DOMAINS.
RX MEDLINE-89181526; PubMed-3070351;
RA Munder T., Mink M., Kuentzel H.;
RT "Domains of the Saccharomyces cerevisiae CDC25 gene controlling
RL mitosis and meiosis.";
RL Mol. Gen. Genet. 214:271-277(1988).
RN (5)
RP FUNCTION.
RX MEDLINE-91203884; PubMed-2017169;
RA Jones S., Vignais M.L., Broach J.R.;
RT "The CDC25 protein of Saccharomyces cerevisiae promotes exchange of
RL guanine nucleotides bound to ras.";
RL Mol. Cell. Biol. 11:2641-2646(1991).
CC -I- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP. THIS
CC PROTEIN POSITIVELY CONTROLS THE LEVEL OF CELLULAR CAMP AT START,
CC THE STAGE AT WHICH THE YEAST CELL DIVISION CYCLE IS TRIGGERED.
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X03579; CAA27259.1; -
DR EMBL, M15458; AAA34478.1; -
DR EMBL, U17247; AAB67360.1; -
DR EMBL, U20618; AAB64528.1; -
DR PIR, A26596; RCBYC5;
DR HSSP, P29355; ISEM;
DR SGD, S0004301; CDC25;
DR InterPro: IPR000651; -
DR InterPro: IPR001452; -
DR InterPro: IPR001895; -
DR Pfam: PF00617; RASGEF_1;
DR Pfam: PF00618; RASGEF_1;
DR Pfam: PF00018; SH3_1;
DR PRINTS: PR00452; SH3DOMAIN;
DR PROSITE: PS00720; GDS-CDC25; 1;
DR PROSITE: PS50002; SH3; 1;
KW Guanine-nucleotide releasing factor; Cell division; Cell cycle;
KW Mitosis; Transmembrane; SH3 domain;
FT TRANSMEM 1452 1473 POTENTIAL.
FT DOMAIN 58 128 SH3.
FT CONFLICT 497 497 I->Y (IN REF. 2).
FT CONFLICT 954 963 PVGHEPEKN->LSVTMNSR (IN REF. 2).
SQ SEQUENCE 1589 AA; 179091 MW; CA90053435C85465 CRC64;

```

RESULT 24
 CC25_YEAST STANDARD; PRT: 1589 AA.
 AC P04821;
 DT 13-AUG-1987 (rel. 05, Created)
 DT 01-JAN-1988 (rel. 06, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 DE CELL DIVISION CONTROL PROTEIN 25
 GN CDC25 OR CTN1 OR YLR310C OR L2142.6
 OS Saccharomyces cerevisiae (Baker's yeast)
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-87131091; PubMed-3545497;
 RA Broek D., Toda T., Michelli T., Levin L., Birchmeier C., Zoller M.,
 RA Powers S., Miglier M.;
 RT "The S. cerevisiae CDC25 gene product regulates the RAS/adenylate

Best Local Similarity	19.9%	Pred. No. 1.3;
Matches	123; Conservative	105; Mismatches
		259; Indels
		131; Gaps
		29;

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: 215035; CA78753.1; -
DR EMBL: X75780; CA53406.1; -
DR EMBL: 228068; CA81905.1; -
DR PIR: B44402; B44402.
DR PIR: S39173; S39173.
DR PIR: S44518; S44518.
DR SGD: 5000151; N0P100.
KW Nuclear Protein; Transport; Repeat.
FT DOMAIN 33 571 G-L-F-G.
SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 2.8%; Score 128; DB 1; Length 959;
Best Local Similarity 20.3%; Pred. No. 1.8;
Matches 143; Conservative 85; Mismatches 257; Indels 218; Gaps 34;

OY 102 NFRPLSTFDCSSKESPIIHOKNGQLSRNNGSMFCRNHAEGSGAISADAFSLQHY 161
D 13 NLSEGSNTSSFGGQOS---OOPNSLFGSNNNNNSTNNASGSGFGFTSAAGSN--SNS 66
OY 162 LPTAFENSSKNGGALIOAFSLSRNVPISFARNRADLNGAICCSNLICS--GNVNP 219
D 67 LF---GNNTNQNNGAF--GQSMGATQN--SPFGLSNNSNAGNTFGSSSMGSGCNTNN 120
OY 220 LFTGNSATN-----GGAICCIDLNTSEKGSISLACNOET----- 255
D 121 AFNNNSNSTNSPFCKNPKMTGCTL--FGSQNNNSAGTSLSFGQSTSTTGTCGTCSSFG 178
OY 256 -----LF--ASNSAKKKGAIYAKHMYLRYN-----GPVSFINNSA-----K 290
D 179 TGLNGNSNIFGAGNNSQNTTGL--FGNOQSSAFGTNNQOGLFGQSOQNTNNAQONQ 238
OY 291 ICGAIALIOGSGSLIAGEGSVLFQNNNSQRTSDQGLVRNAYILEKDALISLEARNQDIL 350
D 239 LQGS---SFGSKPV--GGSLFGQSN-----NTLGNNTNNRNG--L 272
OY 351 FPDPIVQESSKESPLPSL--QASVTSPTPA---TASPLVYIOTNSANSVIFSSRLSE 404
D 273 F---GOMNSNQGSSNSGLFGQNSNNSSTQGVFGQNNQMOQINGNNNSLFGKANTFSN 328
OY 405 EKKTPNLTLSQLOPELKSRLVLRVLAVALSPSLDOPALLINEAGTSLSKTSDDLK 464
D 329 S-----ASGGLFGQNNQOQSGGLFGQNSOT-----SGSS----- 357
OY 465 ATLSPLSLHSDTEKSVYTHAPNLISQIFLNSGDNFYNVELLSKEO----- 513
D 358 -----GLFGQNNQKQ-----PNTFTQ-----SNTGILFGQN-----NNQOQOSTGLFGAKP 399
OY 514 -----NNIPLLLSKEOSH-----LHLP--DGNLSHFGY 542
D 400 AGTTGSLFGQNSSTQPNLSFGITNVPNTSNTQSOQNSLFGCATKLTMMPCGNPTJANQSGS 459
OY 543 GDMTFMKSDSGEHLIANTPKNYVP-----HPEROSTLVANT--LWNTYSDMOAVOS 594
D 460 GNLFGCTKPASTTGLSGFNNTASTYTPSTNGILFGNANNANSTSTTNGILFGAKPDQSKPA 519
OY 595 MINTIAHGCAVLEFGTQGSVAVALFYAHDSGKPDJMMHRSIAGLYLFGISTHSLDHSFCL 654
D 520 L-----GGGLFGQNSNSSTI-----GQNKPVFGGTTQNTGLFGATGINS-----SAVG 563
OY 655 AAOQLLQKSSDSFTTSTETTSYI--ATVQQLATSLMKISAQA 695
D 564 STGKLFQGNNTLNVTGONVPPVNTTQNALCTTAVPSLOQA 606

RESULT 19
YEF3_YEAST
ID YEF3_YEAST STANDARD: PRT: 956 AA.

AC P32618;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 106.1 KDA PROTEIN IN GLY1-GD1 INTERGENIC REGION.
CN YEL043W OR SYGP-ORF14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
RA Wei Y., Taylor P., Nakahara K., Roberts D., Davis R.W.,
RL Submitted (FEB-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hummel-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petrel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.,
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP FIBRONECTIN TYPE III DOMAIN.
RX MEDLINE=97148176; PubMed=8994808;
RA Bateman A., Chothia C.;
RT "Fibronectin type III domains in yeast detected by a hidden Markov
RT model."
RT Curr. Biol. 6:1544-1546(1996).
CC -I- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U18779; AAB64999.1; -
DR PIR: S30834; S30834.
DR SGD: 50000769; YEL043W.
DR InterPro: IPR001777; -
DR Pfam: PF00041; fn3.1.
KW Hypothetical protein.
FT DOMAIN 35 125 FIBRONECTIN TYPE-III
FT DOMAIN
SQ SEQUENCE 956 AA; 106132 MW; 3F7809A0FCA03AF CRC64;

Query Match 2.8%; Score 127.5; DB 1; Length 956;
Best Local Similarity 20.7%; Pred. No. 1.9;
Matches 132; Conservative 91; Mismatches 227; Indels 187; Gaps 31;

OY 88 KEALFK--EKGLDLSNFRPLSTFDCSSKESPIIHOKNGQL-----SLRNGSM 136
D 330 KEKFFKLINESNASVANIN---KEIESLQNEISKMEESKRLNASKSLITSIYNAVY 385
OY 137 SFCRNHAEGSGAI--SAQDFSLQHN-YLPTAFENSSKNGCAI----- 178
D 386 ENDKPIASGSLAVLKLKLNDFLEKNGFLSNAGEEFLSKLNADSSLSLKMQLSDQEL 445
OY 179 ---OQOTFSLSRNVPISFARNRADLNGAICCSNLICSGVNPPLFTGNS--AATN--- 229
D 446 EAMWKLDORSNLKKAISALEQNFMSLNN-----RNLTKLTVQPKYKNNQDSLAATNSNN 500
OY 230 -----GGAICCIDLNTSEKGSLSLACNOETLFSNSAKKKGAIYAKHMYLRYNPV 282
D 501 SAEKNRSSGSIQLPLSNMMSRTGSDIDLISNNKRSINNSNADSA-----PLRLHNPV 552

RP ALTERNATIVE SPLICING.
RX MEDLINE:94061824: Pubmed-8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Takeo M.;
RT "APC gene messenger RNA: novel isoforms that lack exon 7."
RL Cancer Res. 53:5589-5591(1993).
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC -1- ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY)
CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATEININS (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1 (SHOWN HERE), 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART, LUNG, BRAIN, STOMACH, TESTIS AND OVARY.
CC -1- PM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC EMBL: M88127; AAB59632.1; -
DR EMBL: U02937; AAA03443.1; -
DR HSSP: 002248; 28CT.
DR MGD: MGI:88039; APC.
DR InterPro: IPR000225; -
DR Pfam: PF00514; Armadillo_seg; 4.
DR PROSITE: PS50176; ARM_REPEAT; 1.
DR Anti-oncogene. Phosphorylation; Alternative splicing; Repeat;
KM Colled coil.
KM DOMAIN 1 61 COILED COIL (POTENTIAL).
FT DOMAIN 125 245 COILED COIL (POTENTIAL).
FT REPEAT 1 728 LEU-RICH.
FT REPEAT 451 493 ARM 1.
FT REPEAT 503 545 ARM 2.
FT REPEAT 546 589 ARM 3.
FT REPEAT 590 636 ARM 4.
FT REPEAT 637 681 ARM 5.
FT REPEAT 682 723 ARM 6.
FT REPEAT 724 765 ARM 7.
FT REPEAT 765 813 SER-RICH.
FT DOMAIN 739 2834 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1356 1375 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 1864 1891 HIGHLY CHARGED.
FT VARSPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARSPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT VARSPLIC 120 120 T -> A (IN STRAIN CAST/EI).
FT VARSPLIC 493 493 V -> I (IN STRAIN CAST/EI).
FT VARSPLIC 797 797 Y -> F (IN STRAIN CAST/EI).
FT VARSPLIC 1330 1330 A -> T (IN STRAIN CAST/EI).
FT VARSPLIC 1618 1618 A -> S (IN STRAIN CAST/EI).
FT VARSPLIC 2294 2294 G -> A (IN STRAIN CAST/EI).
FT VARSPLIC 2496 2496 H -> Q (IN STRAIN CAST/EI).
FT VARSPLIC 2523 2523 T -> A (IN STRAIN CAST/EI).
FT VARSPLIC 2813 2813 T -> S (IN STRAIN CAST/EI).
FT SEQUENCE 2845 AA: 311086 MW: 145CAV73CF570A499 CRC64;

Query Match 2.9%: Score 130.5; DB 1; Length 2845;
Best Local Similarity 19.0%: Pred. No. 6.5; 310; Indels 355; Gaps 37;
Matches 177; Conservative 92; Mismatches 310; Indels 355; Gaps 37;

OY 65 DPLY---VGNSTYCW---FVSKLHTIDPKKALEKKGDLSTIONFRPLFTDCSSKSS 117
DB 1839 DSHHHYPIPIGTGYCFGRNDLSLDDDDVDLSREKAELEL-----KGKES- 1885
OY 118 PSIIHOKNGOLSLRNNGSMFCRNHAGCGAISADAFSLQHN-----YLFTAFE 167

DB 1886 -----KDEFAKVTCPREPNSOQASQASIKHPANRAQSPVLQKQFTFP 1932
OY 168 ENSSKG--NGCAIQAFSTLSRNVSPISFANRADLNGAICGNCNLICGVN----- 218
DB 1933 QSKDGPORCAATDEKILONLAIENTPVCFSR-----NSSLSLSDIDQENNNNESEPIK 1987
OY 219 -----PLFETGNSATNGAIC-----CISDLNT 241
DB 1988 EAEPANSOGEPSKQASQVAPKSFHVEDTPVCFSSNSSLSSLSIDSEDDLQECISSAMP 2047
OY 242 SEKGSLSLACNOETLFPASNSAKEKGAIYAKHMYLRY----- 278
DB 2048 KKKRPSRLKSSSE-----KQSPRVGGILADLTLDLKDLPQSEHAFSPGSENFOWKA 2102
OY 279 -----NGPVSTFNSAKIGAIQAIQ-SCGSLSTIAGESSVLFONNSORTSOG-----LVR 328
DB 2103 IQEGANSIVSSLHQAAMAAALSRQASSDSSTLSLKGISLGSFHLTPQOEKPTSN 2162
OY 329 NAIYLEKAILLSLEAR----- 345
DB 2163 KGPRLKGEKSTLEAKIEBENKGIKGGKYYKSLITKIRSNSEISSOKKQPLPTMP 2222
OY 346 ---NGDILFPPIVOESSKESPL-----PSSLQASVTSP---TA---TA 382
DB 2223 SISRGRTMIHPIGLRNSSSSTSPVSKKGPPLKTPASKSPSEGCATTSPPRGTKPAGKSEL 2282
OY 383 SPLVIQTSANRSVITFSERLSEEEKTPDNLTSQ-LQOPELKSGLVLKDRVILSAPS-L 440
DB 2283 SPITROTQISGSKNGKSSRSGSRDSTPSRPTQOPLSRPQSGRNSISPRGRGISPPKML 2342
OY 441 SODPOA-----LLIMEAGTSLKTSVDIKLATISPLHSLDTEKSVT----- 481
DB 2343 SOLPRTSSPSTASTKSSGSKMSTSPGKQLSQONLTKQASLSKNASTPRSEASKGLN 2402
OY 482 -IHAENLSIOKIFLS-----NSGDEFYENVLLSKQENNI-----PLLTSKEOS--- 526
DB 2403 QMSNGSKKKVYELMSSTKSSGSESDSERPALVROSTFLKEAPSLRKRKLEESASF 2462
OY 527 -----HLHLPQNLSSIHQYQ-GDW-----TFESKWDs--- 552
DB 2463 ESLSPSSRPDSPTRSQAQTPVLSPLDMSLSTHPSVQAGWRKLPPLNLSPTTEYDGRP 2522
OY 553 DGGHSLIAMMTKKNVPHPEROSTLVANTLWNTYSDMAVQSMINTIAGCAVLFGTWGS 612
DB 2523 TKRHIDARS-----HSESPRLPINRA-----GTMKR 2549
OY 613 AVSNLFYAHDSGKPLDNMHHRSGLYLFGIETHSLDHSFCLAAGQLLCKSSDSFITSTE 672
DB 2550 EHSK-----HSSSLPRVSTW---RRTG-----SSSSLSASS 2578
OY 673 TTSYIATVOQLATSLMKLSAQACYNESIHETKTKRPSKEGFGSMHVSAN-----GE 727
DB 2579 ESSEKAKSEDERIVSSMPAPRQKRN-----QVPTK-----GTWRKIKESDISPTGM 2625
OY 728 VCASTIVSNG--SGLFSSPSIFSKLOGFSGTOD 759
DB 2626 ASOS---ASSGAASGAESKPLITQMAPPVSKTED 2656

RESULT 17
BEM3_YEAST
ID BEM3_YEAST STANDARD; PRT: 1128 AA.
AC P32873;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE BEM3 PROTEIN.
GN BEM3 OR YPLI15C OR LPH12C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]

[illegible]

01-NOV-1995 (Rel. 32, Created)
 01-NOV-1995 (Rel. 32, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 HANSENULA MRAKII KILLER TOXIN-RESISTANT PROTEIN 1 PRECURSOR.
 CN HKR1 OR YDR420M.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YN 295;
 RX MEDLINE=94156857; PubMed=8113191;
 RA Kasahara S., Yamada H., Mio T., Shiratori Y., Miyamoto C.,
 RA Yabe T., Nakajima T., Ichishima E., Furutachi Y.;
 RT Cloning of the Saccharomyces cerevisiae gene whose overexpression
 RT overcomes the effects of HM-1 killer toxin, which inhibits
 RT beta-glucan synthesis.;
 RL J. Bacteriol. 176:1488-1499(1994).
 CC - FUNCTION: COULD REGULATE BETA-GLUCAN SYNTHESIS. OVEREXPRESSION
 CC PROVIDES RESISTANCE TO HM-1 KILLER TOXIN.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (PROBABLE).
 CC - PTM: COULD BE O-GLYCOSYLATED IN SERINE/THREONINE RICH DOMAIN.
 CC - SIMILARITY: SOME, TO YEAST MSB2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: S69101; AAB30051.1;
 DR SCD: S0002828; HKR1.
 KW Glycoprotein; Transmembrane; Repeat; Signal.
 KM
 FT SIGNAL 1 21
 FT CHAIN 2 1802
 FT TRANSMEM 1486 1506
 FT DOMAIN 23 1478
 FT DOMAIN 453 788
 FT
 FT REPEAT 453 480
 FT REPEAT 481 508
 FT REPEAT 509 536
 FT REPEAT 537 564
 FT REPEAT 565 592
 FT REPEAT 593 620
 FT REPEAT 621 648
 FT REPEAT 649 676
 FT REPEAT 677 704
 FT REPEAT 705 732
 FT REPEAT 733 760
 FT REPEAT 761 788
 FT CARBOHYD 24
 FT CARBOHYD 1252 1252
 FT CARBOHYD 1293 1293
 FT CARBOHYD 1342 1342
 FT CARBOHYD 1400 1400
 FT SEQUENCE 1802 AA: 188890 MW: E344CA6469785A24 CRC64;
 Query Match 3.0%; Score 135; DB 1; Length 1802;
 Best Local Similarity 19.5%; Pred. No. 1.7; 361; Indels 164; Gaps 29;
 Matches 158; Conservative 128; Mismatches 164; Gaps 29;
 14 AIIISTAVLFGDPLGFTALITKPNHNV---VCTFEDCTMESLF-----PALCAH 61
 375 AVSISITNTIVSASSTVSTSTNTPVSVSHVATSSSTYISSLSVSPSLAS 434
 62 ASODDPLIVLGNICYCF-----VSKLHTDPKALFKKGGDLSTQNFRLSTDCSSKE 115

435 VSHFGVAPFPFSATISFSSVPVAVASTYTSSPSASVVP-----SAXA 477
 116 SPSIITHOKNQLSLRNNGSMSCFCRNHAEAGGCAISADAFSLQHNLYLPTAEENSSKNG 175
 478 SSPSV-----PVAVSSTYTSSPSAPAAIISSTYTSSPSAPVAVSSTYTSSPSAP 525
 176 GAIQOTSLSHNNPSPISFARRARALNGAICCSNLICSGVNNPLFTFGNSATNGAICCC 235
 526 AAI-SSTYTSSPS-APVAVSSTYTSSPSAPAAIISSTYTSSPSAPVAVSSTYTSSPSAPAA 583
 236 ISDLTSEKGSLSLACNQTETLFASNAKEKGAIAKHMVLRNPGVFINNSAKIGGAI 295
 584 ISSTYTSSP-SAPVAVS-STYTSSPSAPAAIISSTYTSSP-----SAPVAVSSTYTSSPSAP 637
 296 AIQSGGSLILAGEGSLVLPQNNSTQSDGLVRNAILYEKALISLRLARNGDLFFDPI 355
 638 AAIIS-----STYTSSPSVPAVASTYTSSP-----SAPAAIISSTYTSSPSV-----PV 680
 356 VOESSKESP-LPSSLQASVTSPTPATAPLYQTSANRSVIFSSERLSEETKTP-----D 410
 681 AVSSTYTSSPSAPAAIISSTYTSSPSAPVAVSSTYTSSPSAPAAIISSTYTSSPSAPVAVS 740
 411 NTSLOQPIELKSGRLVLDKRAVLAPSLSDPOLLMENAGTSLKTSDDLKATLSIP 470
 741 YTTSPSAPAAIISSTYTSSPSAPVAVSSTYTSSPSALVY-----STSTSSPYDI----- 791
 471 LHSLDTEKSVTI-HAPNLSIQIFLNSGDNFENVELLSKEQNNIPLLLSKEQS-HL 528
 792 VSPSTFAIISGGYTPSPS-ASVAMSTSSSPYDIVLSASSASRSIATYEFSPSPST 850
 529 HLPDGLSHFPGYQGWITSMKDDSGHSLIANMTPKNVPIPERSTLYANTLNNITSD 588
 851 SLPTSTYTFSS-----AAAFESSERYSTSTIAF-----TQIHSTL-SRTID 894
 589 MOAVOSMINTIAHGAYLFGTWSAVSNLFYAHDSGKPIDMWHRSLEYLFGISTHSD 648
 895 FLLOTSM-----AIGSIVSOQISSTSLNDEIHSSALSYNPASNLV 937
 649 DHSFCLAGLOLKSSDSFTTSTETTSYATVOALATSLMKISAQACYNESIHELKTY 708
 938 ETS-----LIISTOASITSPKNSAKISLSLOSLSS-----TKNPY 974
 709 RPSKSGCFSGMSHVAVSGEVCAS-----IPIVNSGSLFSSRSIFSKIO 752
 975 DTANKNTETSGRSTVSNFLYTSSAAKPNKEKFSATPTETITSSSHAY-SLSTPSSH 1033
 1034 SVTGLSHNFVDSKSKATSFQYSSSSISIKL 1064
 RESULT 13
 OMPB.RICPR STANDARD: PRT: 1643 AA.
 AC Q53020; Q9ZCWO;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
 DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
 DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)
 DE (120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].
 GN OMPB OR SPAP OR SPA OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BREINL;
 RX MEDLINE=91045972; PubMed=2122457;
 RA Carl M., Dobson M.E., Ching W.M., Dasch G.A.;

01-JUN-1994 (Rel. 29, last sequence update)
 01-JUN-1994 (Rel. 29, last annotation update)
 HYPOTHEICAL 122.2 KDA PROTEIN IN SIR1 3'REGION PRECURSOR.
 DE YKRI02W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NX NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Gallion L., Dujon B.:
 RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN FLO1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 228327; CAA82182.1;
 DR PIR: S38181; S38181.
 DR SGD: S0001810; YKRI02W.
 KW Hypothetical protein; Glycoprotein; Repeat; Cell wall; Signal.
 FT SIGNAL 1 26
 FT CHAIN 1 26
 FT DOMAIN 27 1169
 FT CARBOHYD 122 122
 FT CARBOHYD 157 157
 FT CARBOHYD 279 279
 FT CARBOHYD 389 389
 FT CARBOHYD 452 452
 FT CARBOHYD 515 515
 FT CARBOHYD 578 578
 FT CARBOHYD 656 656
 FT CARBOHYD 686 686
 FT CARBOHYD 879 879
 FT CARBOHYD 1092 1092
 FT CARBOHYD 1099 1099
 FT SEQUENCE 1169 AA; 122164 MW; 940C6DFB3569C669 CRC64;
 SQ
 Query Match 3.2% Score 143; DB 1; Length 1169;
 Best Local Similarity 17.7% Pred. No. 0.27; Indels 224; Gaps 26;
 Matches 131; Conservative 119; Mismatches 267;
 DB 109 TDCSSKSSPSIIHOKNGQLSLRN-----NGSMFCRNHAEGSGAISADAFSLQHYLP 163
 DB 555 TECTETSTSTVYPPYSSSTAAANYTSSFSSESEVCTECTETSTST-PRYATSTGTA 613
 DB 164 TAFEENSSKNGCAIQAT---FSLSRNVSPISPARNRADLNGAICCSNLICSGANNPL 220
 DB 614 TSFTASNTMTSLVQTDITVYFSLSTVSHT----- 646
 DB 221 FTFGNSATNGCAICISDLNLTSEKGLSLACNOETLFASSAKKEGGAIVAKHVLRYNG 280
 DB 647 -----NAFTS-----SVESNASTFTISSKNGSKSVYTTSSIHSTTPYRP 684
 DB 281 PVSFINNASKIGAI---AIOGGSLSLAGEGVLFQNNQORTSDGLVNAIYLEKDA 337
 DB 685 SNQVITSSSVYPTITSSSSSSASVTLIPETITSEFKPTMKTK----- 729
 DB 338 ILSLEARNGLIFPDPIVQSSSKESPLPSLOASVSPRPATAS-----PLVIOISA 391
 DB 730 VVSISSTPTNLITSDYDTKSDTSSSVSLSSISLPSYSASSQOIFHSSIVSSNG 789
 DB 392 NRSVIFSEERLSEEEKPTDNLTSLOQPIELKSGRLVLDRAVLASAPLSODPALLIME 451
 DB 790 QALTFSSTKVSSESSSSSHRTS-----PTTSSESG---IKS 823

452 AGTSUKTSSDCLKATLSPHSLDT-EKSVTHA-----PNLSIOKIFLNSGDENFENV 506
 DB 824 SVEIEESTS-----TSSSFHETSTASTVSQISOFTVPSSPISTVAPRSTGLNSOTEST 878
 DB 507 ELLSKKONNIPLLTTLKSKOSHLPLPDGNLS-----HFGYQGMWTFMSWMDSEG 555
 DB 879 N-SSKE-----TSSSEASASVMSSSATSPKTKQKYNDSNDSGFSRDRITVYMTSE- 929
 DB 556 HSLIANMTPKNVYPPPEROSTLV-----ANTLMNTYSDMAVQSNINTIAHGAYLGT 609
 DB 930 -----TPSTNEQTLITVSSCESNCSNTVS--SAVSTATTINGITTEVTT 975
 DB 610 KGSANSLNFYAADSSKPLDMMHHRSLGLVLFQISTHSLDDHFC-LAAGQLLGKS---SD 665
 DB 976 W-----CPLSATVETLVSKLESE 993
 DB 666 SFTSTETSTYATVOALATSLMKISAQACYNESIHELTKYRSFKSGFSGMSHVAVS 725
 DB 994 EKTLLITVTSCEGSCGSETASPAIVSTATATVNDV-----TVYSTWSP---ATNKLAYS 1046
 DB 726 GEV-----CASIPVNSGSLFSSFSIFSKLOGFSGTODGFEESSGEIRSESA 773
 DB 1047 SDLENSAKSAFVSEAAETKRSISRNNN-FVPTSGTTSIEHTHTTTSNASENSDNVSASEA 1105
 DB 774 SSFRNLSPLPGITFEKKSOKT 794
 DB 1106 VSSKSVTNPLVLSVSQOPROT 1126
 RESULT 10
 Y144_HUMAN
 ID Y144_HUMAN STANDARD: PRT: 983 AA.
 AC 014157;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, last sequence update)
 DT 01-OCT-2000 (Rel. 40, last annotation update)
 DE HYPOTHEICAL PROTEIN KIA0144.
 GN KIA0144.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 CC NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RP TISSUE-Bone marrow;
 RC MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT The coding sequences of 40 new genes (KIA0121-KIA0160) deduced by
 RT DNA Res. 2:167-174(1995)."
 RL
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D63478; BAA09765.1;
 DR InterPro: IPR000449;
 DR Pfam: PF00627; UBA: 1.
 KW Hypothetical protein.
 SQ SEQUENCE 983 AA; 103930 MW; 50579CFA293EBA68 CRC64;
 Query Match 3.1% Score 139.5; DB 1; Length 983;
 Best Local Similarity 23.0% Pred. No. 0.35; Indels 183; Gaps 30;
 Matches 141; Conservative 64; Mismatches 226;
 DB 237 SDLNTESEKGLSLACNOETLFASNS-----AKERGAIAIAKHVLYRYNGPYVFINNSAKI 291

or send an email to license@isb-sib.ch.

CC -----
 DR EMBL: M77354: AAA34798.1;
 DR EMBL: 272799: CAA96997.1;
 DR PIR: S25370: S25370.
 DR SCD: S0003246: MSB2.
 KW Transmembrane: Glycoprotein; Repeat.
 FT DOMAIN 698 816 7 x 17 AA TANDEM REPEATS.
 FT REPEAT 698 816 1.
 FT REPEAT 715 731 2.
 FT REPEAT 732 748 3.
 FT REPEAT 749 765 4.
 FT REPEAT 766 782 5.
 FT REPEAT 783 799 6.
 FT REPEAT 800 816 7.
 SO SEQUENCE 1306 AA; 133114 MW; 675D5D984D5CA4A6D CRC64;

Query Match 3.3%; Score 149.5; DB 1; Length 1306;
 Best Local Similarity 21.1%; Pred. No. 0.12;
 Matches 161; Conservative 106; Mismatches 347; Indels 149; Gaps 33;

OY 107 STTDCSSKSSSIHQ-----KNGQLSLNNGSMSCRNIAEGSGAISADAFSLQH 159
 DB 496 SSSDVAS-TSAPSVVSSSFTYSLAQGSSMTNPSSSTIVYSSSTGSEBSASTAS--- 551
 OY 160 NYLFAFEENSSKNGCAIGAOFTSLRNVPISFARNRADLNGAICCNLCGNNP 219
 DB 552 ----ATLSSSTTYMAGNIOQDPSTSLLS-----ESQATSTAVLASSV---STSP 599
 OY 220 IFFTGSAATNGAICCTISDLNTEKG-SLSLACNOETLPASNS---AKEGCAIYKKNV 215
 DB 600 YTAGAGATSEASLISSTSAETSOVSQSTALQTSFSSASTEGSESTSSGFEITSSV 659
 OY 276 LRYNGVSTIN-----NSAKIGCAIAIQSGSILLAGESVLEPQNNSORTSDQ 324
 DB 660 L-VQWSSSISSEFSPQTTOMNSASSSSOYTTISSTGLSQVS-DTSVSTTTSSSVSOV 717
 OY 325 GAVRNAIYLEKDAIISLEARNGLDIFDPRIVOESSKESPELPSLOAVTSPT---AT 381
 DB 718 SDTPVSTTTSSSVSOVDTPVSTTTSSSVSOVDTPVSTTTSSSVSOVDTPVSTYT 777
 OY 382 ASPLVITOTSANRSVIFSSER--LSEBEKTP-----DNLTSQLQPIELKSGRLVK 430
 DB 778 SSSSVSOVS-DTSVPTSSRSRSVSOVDTPVSTSSRSVSQTSSTSLQPTTSSORTIS 836
 OY 431 DRAVLS-APSLQDPOLLMEAGT-----SLKTSIDLKATLSTI-----PLHSLDEKS 479
 DB 837 THGALSSSSSVSOQASEITSSINATASEVHSIOTTAATOSTTISFTJANSSASAPLEVA 896
 OY 480 VTIHAPLSTOKIFLNSGDEFENELLSEKQNNIPILT-LSKEQSHLHPDGNLSH 538
 DB 897 TSTPTPSKASLLTPS-----TSSLQVATNTNVQTSITESTYLEPSTNSS 948
 OY 539 FGYOGDWTFSWKOSDECHSLANMTPKNVYPHPRQSTLVANTLMNYSQMAQSMINT 598
 DB 949 -----TFSLVTSDDNMW-----WIPTELITQAEASTASTIVGAT-----QTM-T 988
 OY 599 IAHGALFTGWSAVNLFVAHDSGKP-IDNMHRSLSLGLFGISTHSLDDHFCILAAG 657
 DB 989 LPH-----AIAAATQVPEDECTLTITICFKALINFEVVSERK-----SSA 1029
 OY 658 QLLKSSDSFT-----STETSYIATVQOLATSLMKISQAQYESIHELK---TK 707
 DB 1030 QIFEYLEALNTPKKNFTNTITVLOIPLDODSLNLYVAVEYFPAETIELSLNLTNS 1089
 OY 708 YRSFSKEGFGSMHSAVAVGECASIPV-----SNQGLFSSTFSIKLQGGSG----- 756
 DB 1090 SSARFYTDOMGTAKMAAM--VDSSIPILGLHDNSNGSGSDSSSSNSNSGSGSGSN 1147
 OY 757 TODFEESGE-----IRSFASAFRISLPILGITFEKKSOX 793
 DB 1148 SNGSVSSSGNSYODACTLEYSSKSNNSV-----TSKSKKK 1185

RESULT 7
 Y109_YEAST
 AC P40442:
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 99.7 KDA PROTEIN IN SDLI 5' REGION PRECURSOR.
 GN Y11696 OR Y19402.07C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horneill T.S., Hunt S., Jagels K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odeil C., Pearson D.,
 RA Rajandream M.A., Riles L., Rowley N., Skellon J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RA Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
 RL
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 246921: CAA87023.1;
 DR SCD: S0001431; Y11696C.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 995
 FT CARBOHYD 28 28
 FT CARBOHYD 35 35
 FT CARBOHYD 468 468
 FT CARBOHYD 664 664
 SO SEQUENCE 995 AA; 99735 MW; F63E287A03F137EC CRC64;

Query Match 3.3%; Score 148.5; DB 1; Length 995;
 Best Local Similarity 19.8%; Pred. No. 0.095;
 Matches 130; Conservative 112; Mismatches 278; Indels 135; Gaps 23;

OY 181 QTSLSLNVPISFARNRADLNGAICCNLCGNNPILFTGNSATNGAICCTISDLN 240
 DB 24 QVYNSNTSISNSSSTSVSSSGSVSISSIAE-----ISSATD--ILSITQSA 73
 OY 241 TSEKGSILACNOETLPASNSAKEKGAIAKHHVLRNGVPFVINSKIGCAIAIQSG 300
 DB 74 SSTSISVSSVSSSVSSSVSSSVSSSS-----VSVSSSVSSSVSSSVSSSVSS 120
 OY 301 GSLSILAGEGVLEPQNNSORTSDQGLVRNAIYLEKDA-----ILSLEARNGLDIFDPRIYQ 357
 DB 121 SSVQASASTSDVSSVSSVSSASDVSSVSSQSSASDVSSVSSQSSASDVSSVS 180
 OY 358 ESSKESPELPSLOAVTSPTPATASPLVITOTSANRSVIFSESLSEBEKTPDOLLTSLQ 417
 DB 181 QASASTSDVSSVSSSVSSASDVSS--VSSQSSASDVSS--VSSQASASTSDVSSVS 236
 OY 418 OPTELKSGRLVAKDRAVLSAP-SLSODPOLLLMEAGISLTSDDLKATLSTIPLHSLDT 476
 DB 237 QASASTSGVSSGSGSVSSASGSSSPFOSTSSASTASGATNSLSITSSASASATA 296
 OY 477 EKSIV-----TIHAPLNSIO-----KIFL-----SNGSDENFYE 504
 DB 297 SNTLSSSDGTIYLPPTTISGDLTLTGKVIATEGVVAAGAKLTLLDGDKYFSADLVYVG 356


```

QY 179 Q-AOTFSLSRNVSPISFARN-----RADLNGAICCC-SNLCGNNVPLEFTGNSATNG 230
D 184 STAFFVSENOSECFELMDNICIOTNTAGKGAITYAGTSNFESNMCD-LFFINNACCAG 242
QY 231 GA-----IC-----CISDLNT-----SEKGSISLACNCT-----LEFS 259
D 243 GAIFSPICSLTGNRGNIYFVNNRCFKNVETASSSEASDGAIRKVTTRLDVTNRCGRIFPSD 302
QY 260 NSAREKGAITYAKHMYLVYNGVPSPFINNSA-KIGGAIAIOSGSLIAGEGVLFQW- 316
D 303 NITKNYGAITYAPVTLVDNGFTYITINNANKKGAITYIDGTSNKISADNHAIFFENMI 362
QY 317 -----NSORTSDGCLVYNAIYLEK-----DAIISLEARNGDILFEDPI-----VOESS 360
D 363 VTNTNAGTSTSNAPPRRMATIVASSGEILLGAGSSON-LIFEDPIEVSNAGVSF 420
QY 361 SKESPLPSLSQASVTPPTATASPLVITOTSANRSYIFESSEERLSEEEKTPDNLTSOLOPI 420
D 421 NKEAD-----QTG-----SVFSGATVNSADFHQRNLQTKTPAPL 455
QY 421 ELKSGRIYKIDRAVLSAPLSQ-----DPOALLMEAGTS----- 455
D 456 TLSNCFELCIEDHQAULTVNRFTOTGVSISLGNCAVLSCYKNGTGDASNSATILKHIGLNL 515
QY 456 ---LKTSSDLKATLISPLHSLDTEKSVTIAHPNLSTQIFLSNGDENFYENVEL---L 509
D 516 SSLKSGAIEPIPLWVE-PTNNSNNTADTAFTSLSDVXLSTLIDVGNSPYSTDLTHAL 574
QY 510 SKEONNIPILLTSKEOSHLHPDGNL-----SHFGYOGDWTFSKDSDE----- 554
D 575 SSQ-----PMLSTIS-EASDNQLOSENI DFSGLNVPYHGWGLTMAKQODPEPASSATL 629
QY 555 -----GSLIAMPTEKKNYVPHPEROSTLVANTLMNTYSMDQAVOSMINT--IANGG 603
D 630 TDPQKANKRHRILLTLWLPAGYVPSPKHRSPLIANTLMGNN--LATESLKNSAELTPSG 687
QY 604 AVLFGTWSAVSNLFYAHDSCKRPIDNMHRSIGYLFISTSHLDHSCFLAAGOLGKS 663
D 688 HRPWGTGGGLGMVYQDPRENNP--GFHNRSSGYSAGMIAG--QTHTESLKFSQTYTKL 743
QY 664 SDSFTSTETTSIATVQAOLATSLMK-----ISAOCYNESIHELKTKRYSFSEKEG 716
D 744 NENVY-AKNVNSKNYSCQCEMLFSLQEGFLYKLVGLYSTGDHCHHFTYOGNLTISO-- 800
QY 717 GSHSVAVSGEVCASIPYVNGS-----GLFSFSIFSKLOGFSGTODFEES 764
D 801 GTRSCOTMGCAVFEDLPKRFSGTHILLTAPFLGALGIYSSLSHFTEGAVP----- 851
QY 765 SGEIRSFSAAS-FRNISLPIGI--TEKKSQKTRTYVYPLGAYIQLDKRVEGCPVYLK 821
D 852 -----RSSTKTPLINVLVPIGVKSGEMNATHRPQAMTVEL-AVOPVLYRDEPGIAOQLLA 906
QY 822 NAVSWDAPMANLDSRAHMERLINO 845
D 907 SKGIWFGSGSPSSRRHMSYKXISOQ 930

```

RESULT 2

```

ID OMPF_CHLTR STANDARD: PRT: 1034 AA.
AC P38008; 084878;
DT 01-OCT-1994 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.
GN PMPF OR CT870.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=813;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RX MEDLINE=99000809; PubMed=9784136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
RN 121
RP SEQUENCE OF 26-35.
RC STRAIN=L2/434/B0;
RA Blin L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
RA Comanducci M., Christensen G., Birkelund S., Vitreou E., Ratti G.,
RA Pallini V.;
RA Submitted (SEP-1994) to the SWISS-PROT data bank.
CC - SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC DR EMBL: AE001360; AAC68468.1; -
KW Outer membrane; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
FT CONFLICT 34 35 RR -> FH (IN REF. 2).
SQ SEQUENCE 1034 AA; 112392 MW; 445FFC435D63A67 CRC64;

```

```

Query Match 9.9%; Score 451; DB 1; Length 1034;
Best Local Similarity 22.3%; Pred. No. 5.8e-21;
Matches 223; Conservative 146; Mismatches 367; Indels 264; Gaps 41;

```

```

QY 63 SDDPPLYVGNISYSCWFSKIAHTDPEALFKKGDLSION--FRPLSTFD-----C 111
D 105 ASDNANMLFCNNVC-----THNKGCGAIRSGCPTRFLNNDVLYFNINISA 149
QY 112 SKSESSPSIITHOKN-----GOLSRNNGSMSFCNHAEGSGAISA-- 154
D 150 GAVYVCTGDHNEKRGALYAFITLTGNKRTLAFINNMGDGGCAISADTOIISTDTYVG 209
QY 155 --PSLOHNVLTAFENSSKNGCAL--QAOTFSLSRNVSPISFARNRDLNGALICCSN 210
D 210 ILEFNHTLNHPYTOAENNARGAICSRDLCSISNNSPIYFNINOGG-KGGAISAR 268
QY 211 LKICGNVNPLEFTGN-----SATNGAICISDLN-TSEKGSISLACNOETLFPASN 260
D 269 CVIDNKKERTIFFNNSSLGSSOSSASNGCAIOTGGFTLRNKKGSF-----YFDSN 320
QY 261 SAREKGAITYAKHMYLVYNGVPSPFINNSAKIGALAIQSGSLIS--LAGEGVLFQWNS 318
D 321 TATHGCAIINGCYIDIRDNGPVYFLNNSAAMGAFLNLSKRSATNYIHGTGDIIVNNNN 380
QY 319 QRTSDGIV--RNAIYLEKAL--LSLEARNGDILFEDPI-----VOESSKESPLPS 368
D 381 VFTLDGNLGRKRLFHNNNEITPYTSLGAKKDRIFYFDLFQWMEVKNKTSNNPSPFT 440
QY 369 SLOASVTPPTATASPLVITOTSANRSYIFESSEERLSEEEKT-----PDNLTSOLOPIELKS 424
D 441 SRNITVNP-----ETFSGCAVAVFYNGMSDSIRFLMKENHYIK--EAPTLKF 488
QY 425 GRVLKRAVLSAPSL--SQDPOALLMEAGTSIKTS--SDKLATLISPLHSL----- 474
D 489 CTALIEDDAELETFNIFPTONPTSLALGSGATLTGVGHCKLNTNGLVLPITILEGKS 548
QY 475 -----DTEKSVTHAPN--LSIQKIFLSNSG-----DENFYENVELSKQNNIP 517
D 549 PCIRIVNPQDKTONTGTGCTPPSSSTSSISTPMIIRNGSLIVDENYESVYDSMDLSRKA 608
QY 518 ILTLTSKEOSHLHPDGNLSS-----HFGYOGDWTFSW----- 549

```


OY 521 LSKESHL-----HLPDGNLSHFYOGDMTFSMK-----DSEGHSLIANWTP 564
 DB 610 IDVDLHLGAPNAPNAPDLTLGNSSKTYOGSWTLOWEPNPPNPNNSMLKASMTK 669
 OY 565 KNYVHPEROSTLVANTLMTNTYSDMAVOAMINTIAGGAYLFTGWSAVSNLFYAHDS 624
 DB 670 TGYNGPFRVASLVSNSLWGSILDVRSASHASIAOASIDGRAYCRGIMISNFFY-HDOD 728
 OY 635 GKPDNNHHRSLGLFGLSTHSLDHSFCLAGOLLGKSSDSFISTE-----TSYAT 679
 DB 729 ALG-OGYRHISGYSIGANSY-FGSSMFGLAETETFGGRKSDVVCGRSDHTCVGSVYST 786
 OY 660 VOALATSLM-KISAQACNESIHELKTRKRSFSGEFGSMHVSAGVSCASIPIVSNG 738
 DB 767 ROLGSGCLFGDAFRASTGFGNQHKTSY-TFAESNVRMDNVCVGEVGLPIMLA 845
 OY 739 SGLF-SSEFSIFSKLOGFSGTODGFESSGEIRSFSSAFRNISLPIGITEFKKSOKRTY 797
 DB 846 SKLYINELRPFVQAEFAVAEHESFTERGQAREFKSGHMLNLSIPVGYFDRCSKHPNK 905
 OY 798 YFFLCAYIIDLKRDVESGPPVLLKNAVSDAMANDRAHYFRLTNORAL-HRLQTLN 856
 DB 906 YSFMGAYICDAVRSISGTETTLSSKETWTTDAFHILRHGVYVRSMYASLTGNIEVYGH 965
 OY 857 VSCVLRGOSHSYSLDICTYRF 878
 DB 966 GKYEYRDASRGYLSIGSKIRF 987

RESULT 22
 O84879 PRELIMINARY: PRT: 1013 AA.
 AC 084879:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE PUTATIVE OUTER MEMBRANE PROTEIN G.
 GN PMPG.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813:
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=D/UM-3/CX:
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lamell C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.,
 RA "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis."
 RT Science 267:754-759(1998).
 RL EMBL: AE001360; AAC68469.1;
 DR EMBL: AE001360; AAC68469.1;
 SQ SEQUENCE 1013 AA: 107366 MW: F0927743C0A651DD CRC64;

Query Match 16.7% Score 758; DB 2; Length 1013;
 Best Local Similarity 26.2% Pred. NO. 1.9e-45;
 Matches 263; Conservative 139; Mismatches 399; Indels 208; Gaps 34;

OY 26 DPLGETA-----LTKNPMHVCTFEEDCTMESLFPALCAHNASODPLVLGNSYCWFS 80
 DB 53 DSGTGVFAAGELTLKLNLSIALPLSCFGLN-----GSFVLGNGHSLTFRE 101
 OY 81 KLHTDPKAL--FKKGLSTONFRFLSFTDCSS-----KSSPS- 119
 DB 102 NRTSTNGAALSDSANSGLFTIEGKELSFNSCNLSLAWLPATTTNGSOTPTTTSTPSN 161
 OY 120 -IIFHOKNGOLSRNNGSMFCRHAEGSGAISADAFSLQHNVLTFAPENSSKNGCGAI 178
 DB 162 GTTYSKT-DLLLNNEKFFYSYLVSGDGAIDAKSLTVQGISKLCVFOENTAOADGAC 220
 OY 179 QAO7-FSLSRNVSPISFARNRADLNGAIC----- 208

DB 221 QVNTSFSAMANEAPIAFIANVAGVRCGIAAODGQCVSSSTEDPVSFRNTAVEF 280
 OY 209 -----SNLIGSNV-----NPLFFIGNSATN-----GGA 233
 DB 261 DGNVAVGGGITYGNVAVFLNNGKTLFLNNVASPYIAAEOPNTNGOASNTSDNYGGGAI 340
 OY 234 CC---ISDLNTSEKSLSLACNOETLFPASNAKEGCAIYAKHMLRPNCPVSPFINSAK 290
 DB 341 FCKNAQAAGSNSSVSFDEGGVYFFESSNVAAGGAIYAKKLVSANCGVQFLGNLAN 400
 OY 291 ICGAIAIAGSGSLSLAAGEVSLFQNNQSORTSDQGLVR-NAIYLEKDAI-----LSSL 342
 DB 401 DCGAIIYLGESGELSLADYGDILFDGNLKRPAKEMAAVNGVTVSSQAISMGGKITTL 460
 OY 343 EARNG-DILFPPIYOESSKESPLPSLSQASVTSPTATASPLYIOTSANRS-VITSS 399
 DB 461 RAKAGHQLFNDPI-EMANGNQPASS-----EPLKINDGEYTGDIYA- 505
 OY 400 ERLSEEEKTPDRLTSQLQPIELKSGRLVKRAVLSAPSLSDPOALLMEAGTSL--- 456
 DB 506 -----NGNSTLYQNTIEGRIYLRKALVSNLSQTGGS-LYMEAGSTLDFV 553
 OY 457 -----KTSDDLKALTLSPHSL-----DTEKSVTI 482
 DB 554 TPQPPOOPPAANQILTLNLSLSSLIANNNAVTPNTPNPAQDSHPAIIGSTTAGSVTI 613
 OY 463 HAPNLSIOKIFLNSGDETFENVELLSKEQN---NIPLLTSLSEOSHLHLPDGNLSH 538.
 DB 614 SGP-----TFEEDL-DDTAVDRIYDMLGSNOKIDVLKQLQLOPSANAPSDTLTGNEMPK 666
 OY 539 FGYGDMTFWS--KDSDEG-HSLIANWTPKNVPPPEROSTLVANTNTYSDMAVOASM 595
 DB 667 YGQGSMLKAMPNTPANNPPTLKATWTKTGYNPPPERASALVPSLWSLIDISALISA 726
 OY 596 INTIAGGAYLFTGWSAVSNLFYAHDSGKPIDNMHHRSLGLFGLSTHSLDHSFCLA 655
 DB 727 IQASVDGRSYCGLWVSGVSNFFY-HDRDALG-OGYRYISGYSIGANSY-FGSSMFGLA 783
 OY 656 AGQLLGKSSDSFTITE-----TSYIATVOALQATSLM-KISAQACNESIHELKTYR 709
 DB 784 FTEVFGRSKDYVYCRSNHACIGSVYLSLKQALCSYLFEGDAFIRASGFGNQHKTSY- 842
 OY 710 SFSKEGFCGMSHVAASCEVACASIPIVSNGSLF-SFSIFSKLOGFSGTODGFESSGGEI 768
 DB 843 TFAESDVAMDNMNCVLGELGVLPIVITPSKLYLDELRFVGAESYADHESFTEGQA 902
 OY 769 RSEFASSEFRNISLPIGITEFKKSOKTRYTYLGLAYIIDLKRDVESGPPVLLKNAVSDA 828
 DB 903 RAFRSCHLMLNLSVPGVKFDRCSSTHPNKSIFPMGAYICDAYRTISGTOTTLSSHQETWTT 962
 OY 829 PMANDSRAMFRLTNORAL-HRLQTLNVCVLGQSHSYSLDICTYRF 878
 DB 963 DAFHILRHGVIVRGSMYASLTNIEYGHGREYRDTSGYLSAGSKYRF 1013

RESULT 23
 O929G5 PRELIMINARY: PRT: 922 AA.
 AC 0929G5:
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY PMP_1 PROTEIN.
 GN PMP_1.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=CWL029;
 RX MEDLINE=9920606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,

```

OY 641 G1STRHSDHSEFLAAGOLGKSSDPTTETSTYATVQAOL-----684
D 575 GYAKTPEDDLFSAFCOLLGKDRKLYVSKNNANITAGSLYQHISTYSANONLLQNTIG 634
OY 685 ATSLMKISAQACYNESIHELTKY-----RSFSKEGFGSMHSAVAGVCASIPIVSN 737
D 635 AEAFLVLAQALTYCHASNDKMTNMTTYAPKRTYAEIKGMDGNDGFCVELGATVIOETE 694
OY 738 GSGLFSSSISIKLOGFGTODGFEF--SSGELRSASASFRISIPITITPEKKSOKRTI 796
D 695 SSLSDMTSPFLKFLVHTHODDEFKENSDDGRFESSNLNLSPIGIKFERFANNDTA 754
OY 797 YYFLGAYTODLKRDEVEGSP-----VVLKNAVSWDAPMANLDSRAVMEFLTNORAL-HR 850
D 755 SYHTAAASPOL---VRNPDCTTSLVSPDSAVWVTKANNLARSAPMLQAGNYLSHN 811
OY 851 LQTLNASCVLRGOSHSYSLDGTTRF 878
D 812 IEIFSQGFELRGSSRTYNVDLGSKIQF 839

RESULT 19
OY 092898 PRELIMINARY: PRT: 936 AA.
AC 092898:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP_7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.,
  "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
  Nat. Genet. 21:385-389(1999).
RT EMBL: AE001627; AAD18589.1;
SQ SEQUENCE 936 AA; 100079 MW; 8881D78A5D3D194EC CRC64;

```

Query Match 17.1% Score 773; DB 2; Length 936;
 Best Local Similarity 28.0%; Pred. No. 1,4e-46;
 Matches 236; Conservative 143; Mismatches 378; Indels 86; Gaps 22;

```

OY 97 DLSIONFRFLSTDCSSKSSPS--IHQKNGOLSLRNNGSMFCRNHAECSGAISADA 154
D 119 NLTFNDFRSLTISCPSLLSTPTGCAKSYGNLSLTGNSQIIFQNSSDNGVYINKN 178
OY 155 FSLQHNLYFTAPENSS-KNGGAIQAOTFSLSRVSP--ISFARNRDLNGAICCSNL 211
D 179 FLSTQSOFASFRRQAFTGKGGVYATGTTIENSQIVSFGSNLAKGSGALYSTD- 237
OY 212 ICSGVN-PLFTFGNSA-----TNGAICCTSDLTSEKGSLSLACNDETFASSAAEK 265
D 238 NCSTIDNFOVIFDGNAAEAQAQGAIC---TTTDR-TVTLGNANLSFTNTALTLY 292
OY 266 GCAIYAKHMYLRNGPVFINNSAKI-----GGAIAQSGSLSLAGEGVLEFQNSQ 319
D 293 GCAISGLKLVTSAGGPTLFQSNISGSSAGGGGAGINIASAGELALSTSGDITF-NMQ 351
OY 330 RTSQGLVNAIYLEKDAILSLARNG-DILFDPDVOESSKESPLSSQLQASVTPT 378
D 352 VTNSTSTIRANINIIDTKVTSIRAAATQSIYFDYDIPNPGTASTDNLNLADANS-- 409
OY 379 PATASPLVIOANSRVFSSERLSEEEKT-PDNLTSOLOPIELKSRVLKDRANVLA 437
D 410 -----EIEYGAIVFSEKELUSPTKAIANVTSTIRPAVAVLARGDLVLRDGVITF 460

```

```

OY 438 PLSODPOALLMEACTSLKT--SSDLKATLSTPLASHDTEKSVYIHA-----PNLSIOKI 492
D 461 KDLTSPGSRILMDGGTTLUSAKEANLSLNGLAVNLSLDGNKKAALKTEADKNSLSGT 520
OY 493 FLNSGDENFYENNELLSKEONNIPLLTLSEOSHLPDNLN-----SHFGYOGD 544
D 521 IALIDTEGSPFENHL--KSASTYPLLELTAGANGTTLGALSTLTLOEPETHYQGN 578
OY 545 WFSKXKDEGHSLIANTPKNYVHPEROSTLVANTLMTYSDMQAVQSMINTAHGA 604
D 579 WQLSVANNTSSKIGSINTRTGYIPSPERKSNPLNSLWGNFIDIRISNOLIETKSGEP 638
OY 605 YLFTGWSAVSNLFFAHDSCKPIQNMHHRSLGYLFGISTSLDHSCLAGOLGSS 664
D 639 FERELMISGINAFYRDSMPTR--HGFRRHISGGYALGITATPAEDQLTFACQLFADR 696
OY 665 DSFITSTETSYIATVQAOLATSLMKIS-----AQACYN 698
D 697 NHITGKNHGYGASLYFHHTEGLPDIANFLMGKATRAPWVLSLSEIOTIPLSPDAKPSYL 756
OY 699 ESHELTKY--RSFSKGFSGSMHSAVAGVCASIPIVNSGGLFSSFSIFSKLOGFGS 756
D 757 HTDNHMKTYTNDNSIIR--GSMRDAFCADLGSALPFIIVSVPYLLKEVEPFVKQYIYA 813
OY 757 TQDGFESSGFRSFSASFRNISLPIGITEFKKSOKTRTYIYFLGAYTODLKRDEVEGSP 816
D 814 HQODYERHAGRAFNKSELINVEIPGIVTERDSSKSGTYDILMTLIDAYRNPCKQ 873
OY 817 VLLKNAVSWDAPMANLDSRAVMEFLTNORALH-FLQTLNASCVLRGOSHSYSLDGT 875
D 874 TSLIASDANMMAYGYNLARQGFSAVAAHFOVYNHMEIFQGFAREVBRSSKNYNTNLGSK 933
OY 876 YRF 878
D 934 FCF 936

```

```

RESULT 20
OY 09J542 PRELIMINARY: PRT: 936 AA.
AC 09J542:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE
  PROTEIN G FAMILY).
GN PMP_7 OR CP0308.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shira T., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
  Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.,
  "Comparison of whole genome sequences of Chlamydia pneumoniae J138
  from Japan and CWL029 from USA.";
  Nucleic Acids Res. 28:2311-2314(2000).
RT [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Umayam L.A., Ullrich T.,
  Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
  Bowman G., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
  McClarty G., Salzberg S.L., Eisen J., Fraser C.M.,
  "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
  trachomatis AR39.";
  Nucleic Acids Res. 28:1397-1406(2000).
RT [3]
RL EMBL: AF002546; BAA98653.1;
DR EMBL: AF002193; AAF38165.1;

```

```

OY 708 YRSFSEKSGSHSVASGEVCASIPVSNCSGL-FSSSIFSKLOGCFSGDQDEESSG 766
DB 824 YTKLPK-GRCSNSNCWGLGELGNLPITLSSRLNKQIIPVKAQVAYATGGIOENTP 882
OY 767 FLSFSSASSFRNISLPICITTFPEKKSOKTRTYVYFLGAYIQDLKRDVEGSPVLLKNAVSM 826
DB 883 EGRIGHGHLNVAAPVGVRFCKNSHNRPDPYITIAVAPDVYRHNHPDODDTLLPINCATW 942
OY 827 DAPMANLDSRAVMEFLTNORALHR-LQTLNVCYLRGOSHYSIDLGTTFE 878
DB 943 TSGNNLTRSTLLVQASSHTSYNDVLEIFGHCGCDIRTSRQYTLIDGSKLRF 995

RESULT 16
P71133 PRELIMINARY: PRT: 846 AA.
AC P71133:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE POMP91A PRECURSOR.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).
DR EMBL; U65943; AAC15923.1;
KM Signal.
FT SIGNAL.
FT CHAIN 1 16 POTENTIAL.
FT SEQUENCE 846 AA; 90834 MM; 4CDC31DC03C2964E CRC64;
SO

Query Match 17.4%; Score 790.5; DB 2; Length 846;
Best Local Similarity 26.2%; Pred. No. 7e-48;
Matches 264; Conservative 141; Mismatches 363; Indels 169; Gaps 35;

```

```

OY 293 GAIATQ-SGGSLSILAGBSVLPFONNSORTSDG---LVRNAIYLEKDAIILSLRARG- 347
DB 232 GAICIKDSGECSTLANCDITFDGKIIITNGSPYTRNSIDLGSGCKFKLAKGEF 291
OY 348 DILFDDPIVOESSKESPELPSLSQASVTSPTATASPLVITQASRVYSFSSERLSEER 407
DB 292 GAFYDPIANTGSGTEIELNK-----TES---DTTYGKIYFSEKESLDEEK 335
OY 408 T-PDNLTSOLOOPTELKSGRLVKDRAVLSPSLSDOPALLIMEAGTSLKTSQD---L 462
DB 336 TVPANLKSIFYKQPLKIGAGSLVKDGVLEAKKITQTKSGTYVMDLGTTLQTPSSSGETI 395
OY 463 KLATSLPIHSL-----DTEKSVTHAPNLISIOKIFLNSGDETFENVELLSKEQNN 515
DB 396 TLTNLDINIASLGGGGTAPAKLATNTASQALSIANVLNT-DSNTYED-PILSASK-S 452
OY 516 IPLLTLSEKSHLHPDGNLS-----SHEGYGDWTFWSKDSDEGHSILA-NMTPKNVY 568
DB 453 FSAITATSSSTVTPPEINLKNYTPPTHGYGQNMVTVMKQSSAOEKTATLTWQGTYS 512
OY 568 PIPHEROSTIVANTLNTYSQAOVQSMINTAHGAYLEGTWGSVSNLEFAHDSSGKPI 628
DB 513 PNPBERGSLVPNTLWGAESDTRAIQNLMDISYNGADYSRGEFVSSLANFLNKSQDTR- 570
OY 629 DMWHRSLGILFGISTHSLDHSFCLAAQOLICKSDSFITSTETSYIATVQAOQ---- 684
DB 571 KRFRIHSAQYALGVYAQTPSDDYCSAFCOLFKKDKDFVSNSTIYAGSLTYOHSIW 630
OY 685 -----ATSLMKISAQACYNESIHELKTRYR-----FSKEGEGSHSVASV 725
DB 631 NTWNTLLQWTLGAEPVLVMAQLTYCHASNKKMTMTNTYTPKNTPEIKQKMDQNDGCF 690
OY 726 GEVCASIPVNSGSLFSSFSIFSKLOGSFSGTQDGEF-SSGEISFSSASSFRNISLPIC 784
DB 691 YEEGAKAPL-ETASLLFDYSPFVKQLVHAHQDDEFKENSDQGRYFESNNLTNLSMGP 749
OY 785 IYFEKKSOKTRTYVYFLGAYIQDLKRDVEGSPVLLKNAVS--WAPAMANLDSRAVMEFL 842
DB 750 VLELFESHKDTASYMLTAYADYRSNPDCTASLLVSTFSAVWTKANNLRHAFILQA 809
OY 843 TNQRLHRLQTLN-VSCYLARGOSHYSIDLGTTFE 878
DB 810 GNYLALTRNTLFSQFELRGSCRYNIDLSKIQF 846

RESULT 17
P71132 PRELIMINARY: PRT: 847 AA.
AC P71132:
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE POMP91A.
OS Chlamydia psittaci.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=96406378; PubMed=8810511;
RA Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;
RT "Molecular cloning and characterization of the genes coding for the
RT highly immunogenic cluster of 90-kilodalton envelope proteins from the
RT Chlamydia psittaci subtype that causes abortion in sheep.";
RL Infect. Immun. 66:1317-1324(1998).

```

```
Db 700 FCHLEFARDKCFIAHNSRTYCGTLFFKHSHTLQOPNRLCRKAFSESALIEKFPREIDL 759
Oy 691 -ISAGACVNESTHELTCKTKRSKKEGFSMHGVAVSGEVCASIP-IVNSGSLFSSFSIF 748
Db 760 ALDVQVSEFHSQNRNETHYLSL-PESEGSWSNCEIAGIGLDPVLSNPHPLFKETFIQ 818
Oy 749 SKLQFSGTODGFEESGEIRSFSSAFSPFNISLPIGITEFKKSOKRTYTYFLGAIQDL 808
Db 819 MKVEWYVYQNSFESSDGRGFSIGRLNLSPVCAKVOGDIGSYTYDLGSEFVSIV 878
Oy 809 KADVESGPVYLKNAVSDAPMANLDSRAVYFRLTNQRLHLQTLNLNSC-----V 860
Db 879 YRNNPOSTALVMSPDMSKTRGCMNLRCQAFLLRGSN-----NYVNSHCELFHYAME 931
Oy 861 LRQSHSYSLDGTTYRF 878
Db 932 LRQSSRNIVDVGTKLRF 949

RESULT 13
Oy 09871 PRELIMINARY: PRT: 772 AA.
AC 09871:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
PMP_3_2.
CN PMP_3_2.
OC Chlamydia pneumoniae (Chlamydia pneumoniae).
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AP002545; BAA98226.1;
SQ SEQUENCE 772 AA; 82931 MW; 484FC56D35801EB CRC64;
```

Query Match 18.5%; Score 837.5; DB 2; Length 772;
Best Local Similarity 31.0%; Pred. No. 2,7e-51; Mismatches 347; Indels 91; Gaps 22;
Matches 249; Conservative 116;

```
Oy 136 MSFCRNHAGSGCAISADAFSLQHNLYLTAFFENSSKNGCAIO-AQTESLRNVPISF 194
Db 1 MLEFSKNFSTNDGAIKAKTLSTLGTMSALFSENTSSKKGAIQTSDALITNGEVSF 60
Oy 195 ARNRADLNGALICCSNLICSGVNPPLFFTGNSAT-----NGAICICISDLTSEK 244
Db 61 SDMTSSDSGALFTTEASVVISNNAKVSFIDNKVTGASSSTTGDMSGAICA---YRTSTD 117
Oy 245 GSLSLACNOLTFASNAKEGAIAKHYLVLYNGVPVSFINNS-----AKIGAIATIOS 299
Db 118 TKYTLGNOMLSENNITSTAGGAIYKLELSSGILTFSRNSVNGGTAAPKGAIAIIE 177
Oy 300 GGSLSLAGEGVLFQNNQSRTSDGLVRNAYILEKDALISLEARNG-DILFEDYVDE 358
Db 178 SGELSLISADSGDVLFGNTVSTTTCGNSSSIDLGTSAKMTALRSAGRAIYFYDDITTG 237
Oy 359 SSS-----KSPLPSSIQASVTSPTPATAPLVYIQTANSVYFSSRLSEEEKTPD 410
Db 238 SSTVTVDVLTAVNTPADSALQYT-----GNIIFTGKISLSETEADS 278
Oy 411 -NTISOLQPIELKSGRLVAKRAVLAPSLSDOPALLMEAGTSIKTSSDKLATLSI 469
Db 279 KNTSKLQPVLTISGSLISLKHGVTQQTQAFQOADSRLMDVGTITLEPADTSTINNLVI 338
Oy 470 PLHSLDTEKSVTH---APNLISQKIFLNSGDENFENVLLSKQNNIPILLTSLS-KE 524
```

```
Db 339 NISSIDGAKKAKIETKATSKNLISCTITLDPGTGFYENHSLRNPDSYDILEAKSGTV 398
Oy 525 QSHLHLPGDNLS--HFGVQGM-TFSKNDSDGHSLLT--WTKRNYPHREQRSTLYA 579
Db 399 TSPAAYPDPIMGEKFFHYGQTMGPVW---GTGASTATFTFMWTKGYIPNPERIGSLVP 455
Oy 580 NTLAMNTSMQVQSMINFIANG--GAYLFTGWSAVSNLFYAHSSGKPIDNMWHRSLG 637
Db 456 NSLMNAFIDISSLHYMETANECLOGDRAF--WCAGLSNF--HKDSTTRGFRHLSSG 511
Oy 638 YLEGISTHSLDHSFCLAAGOLLGKSSDSFTTSTETSYIATVQALATSLMKI----- 691
Db 512 YVIGLNLHTCSDKILSAACQLFGRDRDYFAKNNQGTVYCGTLTYOHNETHYSLPCKLAP 571
Oy 692 -----SAQACVNESTHELTCKTKYRSFSEEGSHSVAVSGEVCASIPYVN 737
Db 572 CSLSYVTEIPVLFSGNLSTYHTDNDLTKYTYTPVK-GSWGNDSPALEFGGRAPICLD 630
Oy 738 GSGLFSSFSIFSKLOGFSGTODGFEESGEIRSFSSAFSPFNISLPIGITEFKKSQ-KTRT 796
Db 631 ESALFEQYMPFKLQFYVAHQEGFKGTAREFSGSLVNLALPIGIFDESDQDQT 690
Oy 797 YVYFLGAYTODLKRDVESGPVYLKNAVSDAPMANLDSRAVYFRLTNQRLH-RLQTL 855
Db 691 YNLTIG-YTVDLVRSNPDCTTLRLISGDSMKTFGTNLARQALVLRAGNHCFNSNFARS 749
Oy 856 NVSCVLRGQSHSYSLDGTTYRF 878
Db 750 QFSFELRGSSRNYNVDLAKYQF 772
```

RESULT 14

```
Oy 092896 PRELIMINARY: PRT: 973 AA.
AC 092896:
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY PMP_13 PROTEIN.
GN PMP_13.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9920606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shitai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AP002546; BAA98660.1;
SQ SEQUENCE 973 AA; 102761 MW; E02A69F611DEBFEZ CRC64;
```

Query Match 18.0%; Score 816.5; DB 2; Length 973;
Best Local Similarity 26.4%; Pred. No. 1.2e-49; Mismatches 404; Indels 139; Gaps 25;
Matches 252; Conservative 158;

```
Oy 37 NPNHYVCTFEEDCTMSLFPALCAHNS-----ODPLVYLGNSYCFVSKLHITDPKEAL 91
Db 49 DPGTLCIFSGDLYIANLDMALISRTSSCSFNRAGALQILKGQVFSFLNIRSSADGAAI 108
```

```

RESULT 10
086163
ID 086163 PRELIMINARY: PRT: 914 AA.
AC 086163;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE OUTER MEMBRANE PROTEIN 5 PRECURSOR (FRAGMENT).
GN OMP5.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC/CWL-029/VR-1310.
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
RL associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ001311; CA04671.1;
KM Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 >914 OUTER MEMBRANE PROTEIN 5.
FT NON_TER 914 914
SQ SEQUENCE 914 AA: 95603 MM: 8BF33BAB680F5E3 CRC64:

Query Match 18.8%; Score 853.5; DB 2; Length 914;
Best Local Similarity 30.8%; Pred. NO. 2.5e-52;
Matches 278; Conservative 137; Mismatches 382; Indels 107; Gaps 30;

OY 12 CAALISSTAVLFGDGPL--GETALLTKNPNHYC----TFFEDCTMESL--FPALCAHAS 63
DB 20 CSTVAAATRENIGPSDSDGCTNTCTYTPKNTTGIDYTLGCDITLQNGDSALTKGCF 79
OY 64 OD-DPLVVLGNSYCMFVSKLITDPKALFEKGLDSTONFRPLFTDCSKE--SSPS 119
DB 80 SOTTELSLPAGKYS--LSFLNKKSAEG--AALSVTTDKNLSTLGSSLTFLAPS 132
OY 120 II-----HKNQGLSLRNGSMSCFRNHAEGSGAISADAFSLQHNVI,FTAFENS 170
DB 133 SVITTPSGKAYKCGDGLTFDNKGTILFKODYCEENGALSTKNLSLKNSTGSISEFGK 192
OY 171 SKG---NGGAIOAO--TFSLSRVSPISFARNRADLNGAL--CCSLICSGVNNP,FEFGN 225
DB 193 SSATKKKGALCATCTGTDITNTATPTLFSNNIAEAGAGALNSTGCTTGNTS-LVFSN 251
OY 226 SAT---NGAICICISDNTSEKGLSLACNOETLFAASNAKEKGATYAKHMYLRY--N 279
DB 232 SVTATAGNGAL-----SGDADVTISNGSVTFSGNQAANAGAIYAKKTLASGGG 303
OY 280 GPVSTFNN-----SAKIGALAIQSGGLSILAGEGSLVLFQNN--QRTSDGLVNAIYL 333
DB 304 GGISFSNNVQGTAGNGAISILAGCGLSABEAGDITFQNNAIYATTPQTKRNSIDI 363
OY 334 EKDALISSLEARN--DILFFPIYQESSKESPLPSLOASTPTPATASPLVIQTSAN 392
DB 364 GSTAATITNRAISGHSIFFYDITANTADSTDTLNLNKAADGN-----STORS 412
OY 393 RSVIFSSERLSEEE--KTDNLTLSQLOPTELKSGRLVLKRAVLISAPLSODPOLIME 451
DB 413 GSIVFSGERLSEDEAKVADNLSTLKQPVTLAGNLVLKRGVTLDTKGTOTAGSSVIMD 472
OY 432 AGTSUKTSS-DLKLTLSTPLHSLDTEKSVTHAP-----NLSTQKIFLSNGGENTYE 504
DB 473 AGTTLKASTEEVTLTGLSTPVDLSLEGKKVYLAASAASKNNALSPILLDNOG--NAYE 530
OY 505 NVELLSKEON-----NIPLTLSKEOSHLPLPDGNSLHFQYGGDTFSWKD-----SDEGH 556
DB 531 NHD-LGKTODPSFVQLSALGTATTTDVAVPVATPTTHYKYGQGTGMKMTWVDOTASTPKTK 589

```

```

OY 557 SLIANTPKNYVPPEROSTLVANTLWNTYSDMOAVOSMINTIAHGAYLFTGWSAVSN 616
DB 590 TATLANTVGLPNERGCLPLVPSNLSMGSEFDIAQGVIERSLALTLCSDDGFMAAGVAN 649
OY 617 LFVADSSGKPIIDNMHRSGLYLFQISTHSLDHSFLCAAGQLKSSDSFTSTETTSY 676
DB 650 -FLDKDKKGEK-RKYRHKSGCYAIGCAQOTCSENLSFAFCOLGSPDXDFLVANHTDTY 707
OY 677 IATVQAOLATSL-----MKISACQVNESIHELKTKRYSKSGFG 717
DB 708 AGATYQIHTDECGSIGCLLDKLPQSMHKPLVLEGQLAYSHVNDLTKTYAT-PEVKG 766
OY 718 SMHSAVAGVEGECASIPYVSNCGSLFSSFSIFSKLOGFSGTODCEESSGEIRSFASFR 777
DB 767 SWGNNAFNMMLGASHSHSPRYLHCEGDTYAPYIKLNTLYIRNDSFSEKTEBRSDSNLF 826
OY 778 NISLPIGITFEKKSOKRTTYTFGLGATYQDLKRDYESSPVVLLKNVSWDAPMANLDSRA 837
DB 827 NLSLPIGYKFEKFSDCNDFSYDLTSLYVPLDIRNDPKCTTALVISGASWETVANNLARO 886
OY 838 YMR 841
DB 887 LQVR 890

RESULT 11
086164
ID 086164 PRELIMINARY: PRT: 928 AA.
AC 086164;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMP4 OR PMP.11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC/CWL-029/VR-1310.
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
RL associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310.
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RT Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
RL of autotransporting pathogenicity factors."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029.
RX MEDLINE=99206066; PubMed=10192388;
RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=J138.
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RT Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RL from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AJ001311; CA04672.1;
DR EMBL: AJ133034; CAB37072.1;
DR EMBL: AE001628; AAD18593.1;
DR EMBL: AP002546; BAA98658.1;

```

RM [12]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39:
RA MEDLINE-20150255: PubMed-10684935:
RX Read T.D., Brinham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Peterson J., Umayam L.A., Ulterback T.,
Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
McClarty G., Salzberg S.L., Eisen J., Fraser C.M.:
RT "Genome sequences of Chlamydia trachomatis Morn and Chlamydia
pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98652.1; -
DR EMBL: AE002193; AAF38166.1; -
DR TIGR: CP0309; -
SO SEQUENCE 1276 AA; 132126 MW; C97FA40662C88E0B CRC64;

Query Match: 19.7%; Score 894.5; DB 2; Length 1276;
Best Local Similarity 30.6%; Pred. No. 5e-55;
Matches 279; Conservative 144; Mismatches 364; Indels 125; Gaps 29;

QY 68 LYVLNSYCHMFPSKLT-----HTDPKEALFKKGLDSTONFFLSFTDCSSKSSPSSTIH 122
DB 389 LYSKNSSLTGTNTLLFSGNKAATGPNSSANQEG---CGAALFLFLESAS-----VS 437
QY 123 OKNGOLSLRNNGSMFCRNHAEAGSADAFSLQHNLYLFTAFENSSKNGAIOAQT 182
DB 438 TKKG-LWIEDNENVSLSGTATATVSGGAIYATKCALHGNTTLT-FGNTAETAGAIYET 495
QY 183 -FSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTGTSATN----- 229
DB 496 EDFTLTSGTGTFTSTNAKTAGALHTKGTSTFN-KALVFSGNSATATATTDTQEGC 554
QY 230 GGAICG-ISDLNTESEKGLSLACNOETLFAANSKE----- 264
DB 555 GGAICNISESDIATK-SLTLTENESLSFINNTAKRSGGCIYAPKCVISGESINFDGNT 613
QY 265 ---KGAIAKHMVLRVNGPVSFINNSAKIGAIQSGSLSLAGESSVLFOUNSOQT 321
DB 614 AETSGAATSKMLSTANGPVSFTNNSGKGGAITYADSGELSLADIDITFSGN--RA 671
QY 322 SDQGLVRNAILKDALISLEARNG-DLFFDPITVOESSS-----KESPLSSLOASVT 375
DB 672 TEGTSTPNSIHLAGAKAKITKLAAPGHITFYFDPTTMEAPAGCITIELYINPVAKATVP 731
QY 376 SPT-----PATASPLYIQTSA--RSVIFSSERL-SEEEKPTDNLTSOLOPIELKSGRL 427
DB 732 PQPKNGPIASVPVAPAPANPTGTIVFSSGKLPSODASIPANTTTILNOKINLAGCNV 791
QY 428 VLKRAVIAAPSLSDOPALLIMEAGTSLKTS-----DLKATLSLPLHSLDTEKSVT 481
DB 792 VLKEGATLOVYSFTQPPSTYFMDAGTILETTTNNNGSIDLKMLSVNLADLKKRMIT 851
QY 482 I-----HAPNLISQIKIFLSNSGDENFYENVELLSKEQNNIPLLTJLSEKSHLPLDGN-- 534
DB 852 IAVNSTSGGLKISGDLKFHNNEGSFYDNPGL--KANLNLPLFDLSTSGTVLDDFNPIP 909
QY 535 ---LSHFGYOGDMTFMSKDDSEGH-SLIANTPKNYVPHPEROSTLVANTIMNTYSDMQ 590
DB 910 SSMAAPPDGYOQSWTLVPRVAGAGKVTLVAEWOALGYTPKPLRATLIPNSLMAWYVNIH 969
QY 591 AVQSMINTIAHGAYLFTWGSVAVSNLFYAHDSGKPLIDNMHRSLSGLYFGISTSLDH 650
DB 970 SIQOGIATAMSDAPSPHGMIGIGNAF--HODKOKENAGFLIRGIIYVGSMTTPOEY 1027
QY 651 SFCLAGOLKSSDSFTTSTETTSYIATVQAQ-----LATSLMK----- 690
DB 1028 TFAVAFSOLFSGSKDYVVDISKQYAGSLCAOSSYVIPRLHSLRRHYSKVLPLPGET 1087
QY 691 ---ISAQACYNSTIHELTKTYRSFKEGFGSHVAAGVEVCASTPIYNSNGGLSSSTSI 747
DB 1088 PVLHGQVSYGRNHNHMTTKLAN-NTQGSQDMDSHFVAVEGSLPVDLNTRYL-TSYSP 1145

QY 748 FSKLOFSGTODGFEESSETRSFSSAFRNISLIGITFEKKSOXKTRYYFLCATYOD 807
DB 1146 YKILQVSYNOKGFOEVADPRIFDASHLVNYSIPMGITFEKHSAKPPSALLTLTGAVD 1205
QY 808 LKRDVSGPVLLKNAVSMDAPMANLDSRAYMFRLTNQ-RALHRLQTLINVCVLRGQSH 866
DB 1206 AVRQ-HPHQTLSTLTGNTGSMSTFATNLSSQAFPAEASGHLKHLGLDCAAGSCGELRSSR 1264
QY 867 STSLDLGTTTYR 878
DB 1265 SYNANGTRYSPF 1276

RESULT 8
ID Q92899 PRELIMINARY; PRT: 1407 AA.
AC Q92899;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G/I FAMILY.
GN PMP-6.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RA MEDLINE-99206606; PubMed-10192388;
RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
R. A. Linger L., Grimwood J., Davis R.W., Stephens R.S.,
"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
DR EMBL: AE001627; AAD18588.1; -
SO SEQUENCE 1407 AA; 144919 MW; 9CF54AA4614D54C7 CRC64;

Query Match: 19.7%; Score 894; DB 2; Length 1407;
Best Local Similarity 31.1%; Pred. No. 6.3e-55;
Matches 270; Conservative 140; Mismatches 343; Indels 116; Gaps 27;

QY 106 LSFDCSSKESPSIIHQKNQOLSLRNNGSMFCRNHAEAGSADAFSLQHNLYLFTA 165
DB 559 LSFLESAS-----VSTRKG-LWIEDNENVSLSGTATATVSGGAIYATKCALHGNTTLT- 609
QY 166 FEENSSKNGAIOAQT--FSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFT 223
DB 610 FDGNTAETAGAIYETEDFTLTGTGTFTSTNAKTAGALHTKGTSTFN-KALVFS 668
QY 224 GNSAN-----GGAICG-ISDLNTESEKGLSLACNOETLFAANSKE----- 264
DB 669 GNSATATATTDTQEGCGGAILCNISESDIATK-SLTLTENESLSFINNTAKRSGGIYA 727
QY 265 -----KGAIAKHMVLRVNGPVSFINNSAKIGAIQSGSL 304
DB 728 PKCVISGESINFDGNTAETSGAATSKMLSTANGPVSFTNNSGKGGAITYADSGLS 787
QY 305 ILAGGSVLFOUNSOQTSDQGLVRNAILKDALISLEARNG-DLFFDPITVOESSS-- 361
DB 788 LEALIDITFSGN--RATEGTSTPNSIHLAGAKITKLAAPGHITFYFDPTTMEAPASG 845
QY 362 ---KESPLSSLOASVTPT---PATASPLYIQTSA--RSVIFSSERL-SEEEKPTD 410
DB 846 GTIELVYNPVAKATVPQPKNGPIASVPVAPAPANPTGTIVFSSGKLPSODASIPA 905
QY 411 NLTSOLOPIELKSGRLVLKRAVIAAPSLSDOPALLIMEAGTSLKTS-----DLKL 464
DB 906 NNTTLNKNINLAGSNVYLKEGATLOVYSFTQPPSTYFMDAGTILETTTNNNGSIDL 965
QY 465 ATLSLPLHSLDTEKSVT---HAPNLISQIKIFLSNSGDENFYENVELLSKEQNNIPLLT 520
DB 966 KNLVNLDAIDGKRMITIAVNSTSGGLKISGDLKFHNNEGSFYDNPGL--KANLNLPLFD 1023

OX NCBI_TaxID=833558;
RN [1]
RP
RC SEQUENCE FROM N.A.
RD STRAIN-J138;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kohara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RL from Japan and CWL029 from USA.";
RN Nucleic Acids Res. 28:2311-2314(2000).
[2]
RP
RC SEQUENCE FROM N.A.
RD STRAIN-AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Ulteback T.,
RA Berry K., Bass S., Linher K., Weisman J., Khouri H., Creven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Debou R., Kolonay J.,
RA McClarty G., Salberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Moyn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL: AP002546; BAA98654.1; -
DR EMBL: AE002193; AAF38164.1; -
TG TRG: CP0307; -
SO SEQUENCE 930 AA: 97639 MW: 4648896761391C09 CRC64:

Query Match	20.3%	Score 919.5;	DB 2;	Length 930;
Best Local Similarity	31.0%;	Pred. No. 5.2e-57;		
Matches 260;	Conservative 150;	Mismatches 343;	Indels 87;	Gaps 26

```

OY 98 LSIQNFRLSF-----TCCSSKESPSIIHQKNOLSLRNNGSSNPFCC-----NHAEGGGA 14
D 119 LTFIFGSLNLSIIARPCTIVVAGSKTSL-----SAGALNLTQDGTILFESONNSNENANGGA 17
OY 150 ISADAFSLQHNVLFTAFEESSKCGNGAI-QAQTFSLSRNVSPISFARNRADLNGAICC 208
D 175 ITAKTLISISGNTSSITFESSAKKLGCIYSAASISGNTGQLVFMNNKCTGGCAL- 233
OY 209 SNLJCSGVN---NPLFFGNSATN-----GGAICISDINTSEKSLSLACQETLFASNS 26
D 233 -GFEASSSITTONSSLEFSGNATPAAGKGAIVC--EKTGETPTLITGKSLJTFEANS 288
OY 262 AKKGGAIYAAHMYLRYGNPVSFINN-----SAKIGGAIAIOSGSLSIAGEGSVLFON 318
D 289 SVTQGAICAIGHGLDLSAAGPTLFSSNRGNTPAAGKGAIAIADSGLSLSLANQDITFLG 348
OY 317 NS-QRTSOGGLVRNAYILEKALLSSLEARNG-DILFFDPPIVOESSKESPELPSLOASV 374
D 349 NTLSTSAPTSTRNAIYIGSSAKTITLRAAGOSIYFVDPLASMTTASDVL----- 400
OY 375 TSPTPATAPSPVIQTOISANRSVIFSSERLS-BEERTPDMLTSLQOPIELKSGRLVLKRA 433
D 401 TINQPDNSPL-----DISCITVFEGEKISADEAKAADNFTSILKQPLALASCTIALGNV 450
OY 434 VLSAPLSQDPQALLIMEAGTSLKTTSD-LKLATLSTPLSLDTEKSVTTHAP-----N 486
D 457 ELDVNGFQTEGSLTLMOPTGFKLADPDEAISLTLYVDSLALTEGNKSVSIEAGANKTIT 518
OY 487 LSIQKIFLNSGDENEFENVELLSKEONNIPLLTSLKESQSLHJPDGNLS-----HF 539
D 517 LTPSLVFDQSSG-NFEVESHINAFQPLVFPFLAATRASDIY-DALLTSPVQTPPEPHY 573
OY 540 GYQGDWTFSMKDSQDEGHSILANMTPKNYVPRPEROSTVLVANTLWNTYSDMQAVOSMINTI 599
D 574 GYQGWMEWTMADSTASTASCTMTWYTTGPNPERRASVAPVPSLASFTDITFTLOOIWTSQ 633
OY 600 AHGGAYLFGTGAASVSNLFYAHDSGKPIQDMWHRSLYEGCISTHSLDHSFCLAGCOL 655
D 634 ANSLYOQGLMASTANFEH-KDQSGTN-QAFRIKRSIGYIYGSAEDPESFISVACOL 691
OY 660 LKGSDDFTITSTETTYIATVQAO-----LATSLMK-----ISAQACYNES 700

```

```

Db      692  FCKKDDLTIVETSNINYLASLYLQHRALGGLPMPSPSITDMKLDPIILNLAQLSYIT  753
Oy      701  IHELTKYRFSRFSKEGGSMHVAVNGEYCASIP1-VSNGSLFSSFSIFSKIQGFGTOD  759
Db      752  KNDMDTIRTSY-PEOGSMTNNSGALTEGSLALYLPEAFEPFGGYEPFLKFAQVYSRQ  810
Oy      760  GFEESSGELFRSSASFPNISLPICITFEKKSQKTRITYYFLGAYIQDLKRDVSGPVIL  819
Db      811  NEKEGGAARAFDDGDLVNCISIPVIRLEKISEDEKNFEISLAYIDGVYKKNPRSTRSL  870
Oy      820  LKNASVMPAPMNLDSRAYMPERLTQNALH-RLDTILNVSCVLGSGHSYSLDGTYYRF  878
Db      871  MWSGASWISLCKNLKARQAFSLASAGSHLTLSPHVELSGEAAELTGLGSHIYINWDCGLYSF  930

```

RESULT	5	
092393		
ID	092393	PRELIMINARY:
AC	092393	PRT: 930 AA.
DT	01-MAY-1999	(TREMBLrel. 10, Created)
DI	01-MAY-1999	(TREMBLrel. 10, Last sequence update)
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)
DE	OUTER MEMBRANE PROTEIN 11 PRECURSOR.	
CN	OMP11 OR PMP_8.	
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).	
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.	
OX	NCBI_taxid=83558;	

RP	(11)	SEQUENCE FROM N.A.		
RP		STRAIN-CMI029:		
RC		MEDLINE=98206606; PubMed=10192388;		
RX		Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA		Ollinger L., Grimwood J., Davis R.W., Stephens R.S.;		
RT		"Comparative genomes of Chlamydia pneumoniae and C. trachomatis";		
RL		Nat. Genet. 21:385-389(1999).		
RN	[2]			
RP		SEQUENCE FROM N.A.		
RC		STRAIN=VR1310;		
RA		Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,		
RA		Christiansen G., Birkelund S.;		
RT		"Chlamydia proteins containing the GGAI-repeat belong to a subfamily		
RT		of autotransporting pathogenicity factors.";		
RL		Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
DR		EMBL; AE001627; AAD18590.1; -		
DR		EMBL; AJ133034; CAB37068.1; -		
KW		Signal.		
FT		SIGNAL		
FT		1	27	POTENTIAL.
FT		CHAIN	28	930
SO		SEQUENCE	930 AA;	97669 MW; 46A9B5E3BB913C4C CRC64;
				OUTER MEMBRANE PROTEIN 11.

Query Match	20.2%	Score 915.5	DB 2	Length 930
Best Local Similarity	30.8%	Pred. No. 1e-56		
Matches 259	Conservative 150	Mismatches 344	Indels 87	Gaps 26
OY	LSIQNRFLSF----	TDCKSSKSSPSIIHOKNGOLSLRNNGSMSCFR----	NHAEGSGCA	149
Db	119 LTFQFSNLSFIAPAGTIVASGKSTL----	SAGNLINTNDNTLIFESQVNSSEANNNGA	174	
OY	150 ISADAFSIQHNLYLPAPEENSSKSGKAI-QOQTSLSRNVSPISEFARRALNGAICG	208		
Db	175 ITTKTSLISGNTSSITFTFNSAKKIGKAIYSSAAMSISGNTQLPFMNNKGTGGAL--	232		
OY	209 SNLIGSGNV-----NLPFFGNSKATN----	GAACICISDLTSSKGSLSLACNOETIFASNS	261	
Db	223 -GFEASSTITONSSLFESGNTATPDAGKGAICY--	EKTGTPLTTITISGNSSLTFAENS	288	
OY	262 AKKEGAIYAAHMYLVRYGVPVSFINN----	SAKIGAIAIIDSQGSLSLTLAGEGSLVFON	316	
Db	289 SVTGGACIACAGGLDLSAAGPTLFSNNRGNTMAAGKGAIAIADSSLSLSANOCGITFCG	348		
OY	317 NS-QRTSDQGLVRNALYILEKDAIISLSLEARNC-DILFPPIVQESSKESPLPSLSQASV	374		

QY	181	OFFSLSRNVSPISFARNADLNGAICCSNLICGGVNPPLFTTGNSTNGAICICISDLN	240
QY	181	OFFSLSRNVSPISFARNADLNGAICCSNLICGGVNPPLFTTGNSTNGAICICISDLN	240
Db	181	OFFSLSRNVSPISFARNADLNGAICCSNLICGGVNPPLFTTGNSTNGAICICISDLN	240
QY	241	TSEKSSLSLACQOEFLFPAFNSNAKEKGGAITYAKHNVLRYRNGVSPFTNNSAKIGGAIIOG	300
Db	241	TSEKSSLSLACQOEFLFPAFNSNAKEKGGAITYAKHNVLRYRNGVSPFTNNSAKIGGAIIOG	300
QY	301	GSLSLLAGEGCVLFPONNSORTSDOGLVRNATYLEKDAI LSLLEARNGDILFPPIVOESS	360
Db	301	GSLSLLAGEGCVLFPONNSORTSDOGLVRNATYLEKDAI LSLLEARNGDILFPPIVOESS	360
QY	361	SKESPSPSLAASVSPRPATASPLVLOTSANRVSIFSSSESLSEEXTPNNTLSQLOQPT	420
Db	361	SKESPSPSLAASVSPRPATASPLVLOTSANRVSIFSSSESLSEEXTPNNTLSQLOQPT	420
QY	421	ELKSGRLVKDRAVLASPLSDODQALLIMAGCSLKTSSDLKATLUSILPHSLDTEKSV	480
Db	421	ELKSGRLVKDRAVLASPLSDODQALLIMAGCSLKTSSDLKATLUSILPHSLDTEKSV	480
QY	481	TIHAANLSIQIKIFLNSGSDENFEYENVELLSKEONNIPDLTILSKEQSHLPLDGNLSHFG	540
Db	481	TIHAANLSIQIKIFLNSGSDENFEYENVELLSKEONNIPDLTILSKEQSHLPLDGNLSHFG	540
QY	541	YOGDWTFSMKSDSGHSLIAMTPEKKNVYRPHEROSTVYANTLVNNTYSQMOVQSMINTTA	600
Db	541	YOGDWTFSMKSDSGHSLIAMTPEKKNVYRPHEROSTVYANTLVNNTYSQMOVQSMINTTA	600
QY	601	HGAYLVFCTMGSAVSNLFYAHDSGKRPIDNMHRSLSGLYEGISTHSLDHSFCLAAGOLL	660
Db	601	HGAYLVFCTMGSAVSNLFYAHDSGKRPIDNMHRSLSGLYEGISTHSLDHSFCLAAGOLL	660
QY	661	GKSSDSPTTSETTSTYATVOAOLATLMLKISAQACYNESIHLELTKRYRFSFSGEGGSMH	720
Db	661	GKSSDSPTTSETTSTYATVOAOLATLMLKISAQACYNESIHLELTKRYRFSFSGEGGSMH	720
QY	721	SVAVSGEVCASIPIVNSGGLFSSFSIFSKLOGFSGTODGFEESSGEIRSFASSPRNTS	780
Db	721	SVAVSGEVCASIPIVNSGGLFSSFSIFSKLOGFSGTODGFEESSGEIRSFASSPRNTS	780
QY	781	LPIGTFEKKSKOKRTYRYFLGAIYIOLKRDVSGRPVLLKNAVSDAPRANLDSRAVYF	840
Db	781	LPIGTFEKKSKOKRTYRYFLGAIYIOLKRDVSGRPVLLKNAVSDAPRANLDSRAVYF	840
QY	841	RLTNORALHRLQTLNLVNSCVLRQGSNHSYSDLCTTYRF	878
Db	841	RLTNORALHRLQTLNLVNSCVLRQGSNHSYSDLCTTYRF	878
RESULT	2		
Q9PL41	PRELIMINARY:	PRT:	867 AA.
AC	Q9PL41		
OC	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	POLYMORPHIC MEMBRANE PROTEIN G FAMILY.		
GN	TC0267.		
OS	Chlamydia muridarum.		
OC	Bacteria: Chlamydiales: Chlamydiaceae: Chlamydia.		
OX	NCBI_TaxID=83560;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-MOPN / NIGG.		
RX	MEDLINE=20150235; PubMed=10684935;		
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.K., Peterson J., Ulfenberg T., Berry K., Baas S.,		
RA	Lilher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,		
RA	Givim M., Nelson W., Deboy R., Kolonay J., McClarty C., Salzberg S.L.,		
RT	Eisen J., Fraser C.M.,		
RT	Genome sequences of Chlamydia trachomatis Mopn and Chlamydia		
RT	pneumoniae AR39."		
TL	Nucleic Acids Res. 28:1397-1406(2000).		

RN [2]
RP SEQUENCE FROM N.A.
RA Read T.D., Brinham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
RA Hickey E.K., Peterson J., Umayam L.A., Utterback T., Berry K.,
RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA Dodson R., Gwin M., Nelson M., Deboy R., Kolonay J., McClarty G.,
RA Salzberg S.L., Eisen J., Fraser C.M.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AE000294; AAF39136.1; -
DR TIGR: TC0267; -
SQ SEQUENCE 867 AA: 95017 MW: 557994185A9E5652 CRC64:

Mon May 7 09:06:43 2001

us-09-677-752-4.rst

Page 24

Use

FEATURES

Source

Location/Qualifiers
1. .785

/clone.lib".cc.reimhardt CC-1690, stress condition I, normalized. lambda zap II" /note-Vector: plasmidscript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Plasmidscript II SK- plasmids were excised from the lambda zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT	352 a	1 c	411 g	17 t	4 others
ORIGIN					

```
alignment_scores:
  Quality: 99.00
  Ratio: 0.853
  Percent Similarity: 52.489
  Length: 221
  Caps: 8
  Percent Identity: 28.507
```

```
alignment_block:
  uc: 00-677-753-4
```

US-09-677-752-4 x BF865549/rev .

Align seg 1/1 to reverse of: BF865549 from: 1 to: 785

[illegible]

```

484  aprOsmLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluA 501
      ||| |||||::: ||||| |||::: |||
247  TCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 198
      ||| |||||::: ||||| |||::: |||
501  snPheTyGluAsnValGlnLeuLeuSerLysGluGlnAsnAsnIlePro 517
      ::::: |||||::: |||
197  CTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 166
      ::::: |||||::: |||
518  Leu...LeuthrLeuSerLysGluGlnSerIshLeuHisLeuProaspG 533
      ||| |||||::: ||||| ||| ||| |||
165  CTCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 116
      ||| |||||::: ||||| ||| ||| |||
533  yAsnLeuSerSer 537
      ::::: ||| |||
115  CTCCTCTCTCTCT 103

seq_name: gb_est76:BF636683

```

seq_name: qb_est76:BE636683

seq.documental_block:	LOCUS	DEFINITION
BEG36683	1183 bp	EST
rocketeller	0.373	Mastigamoeba balamuchi lambda ZAP II library
Mastigamoeba balamuchi		cDNA similar to glyceraldehyde-3-phosphate dehydrogenase, mRNA sequence.
nc014404		

ACCESSION	BE636683
VERSION	BE636683.1
	GI:9919794

KEYWORDS EST.

SOURCE	Mastigamoeba	balamuthi..
ORGANISM	Mastigamoeba	balamuthi

ORGANISM

REFERENCE
1 Eukaryota; Mastigamoeba
(bases 1 to 1183)

AUTHORS
Lee, J. A., Moore, D. V., Gordon, P., Sensen, C. W., Gaasterland, T. and Muller, M.

TITL

JOURNAL
Unpublished (2000)

COMMENT Contact: Muller M

Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmueller@rockvax.rockefeller.edu
Insert Length: 1183 Std Error: 0.00
POLYA-No.

FEATURES	Location/Qualifiers
source	1; .1183

```

/organism="Mastigamoeba balamuthi"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II library"
/note=syn: Phreatamoeba balamuthi

```

BASE COUNT	212 a	456 c	324 g	191 t
ORIGIN				

```

alignment_scores:
  Quality: 99.00      Length: 361
  Ratio: 0.576      Gaps: 13
  Percent Similarity: 47.645      Percent Identity: 20.222

```

```
alignment_block:
```

US-09-677-752-4 X BE636683 .

Align seq 1/1 to: BE636683 from: 1 to: 1183

[illegible]

```

seq.documentation.block:
LOCUS      A2548467       908 bp    DNA
DEFINITION ENTREC30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  A2548467
VERSION    A2548467
KEYWORDS   A2548467.1 GI:11172102
SOURCE     Entamoeba histolytica.
            Entamoeba histolytica.
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 908)
            Loftus,B., Van Aken,S. and Fraser,C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
            Unpublished (2000)
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjoftus@etigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 828.
            Location/Qualifiers
                1..908
                /organism="Entamoeba histolytica"
                /strain="HMI:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: pHS01; Site_1: Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G. and Diamond, I.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                light size distribution (~2 kb). The v + l method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Borell, Oxford University Press, 1999)."
```

seq_name: gb_test88:BF495834

LOCUS
DEFINITION A090496.Sprine AT Drosophila melanogaster adult testes por7
Drosophila melanogaster CDNA clone A090496.5 similar to X64412:
DKORBR_2 orb FBgn0004882 SPTREMBL:Q24433 X64412:783..3530, mRNA
sequence.
ACCESSION BF495834
VERSION BF495834.1 GI:11579135
KEYWORDS EST.
SOURCE fruit fly.

seq_documentation_block: EST 06-DEC-2000

157 LeuglnHisasnIyLeuPheThrAlaPheGluAsnSerSeryl 173
157 LeuglnHisasnIyLeuPheThrAlaPheGluAsnSerSeryl 173
656 TCTTAAATCATATTTCGTCTTCATCATCATCTTCTTTCTTCATC 607
173 yAsnGlyAlaAlaIleGlnAlaGlnThrPheSerLeuSerAArgasValS 190
606 ATTCTTCATCATCTTCTTCTTCTTCATCACTCTTCTTCATCATCTT 557
190 ePrProIeSerPheAlaArgasnaArgAlaAspLeuasnGlyAlaIle 206
556 CTTCCTTCATCATCATCTTCTTCTTCTTCATCACTTCATCATCTTCT 507
207 CysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheTh 223
506 TCCTCATCATCATCTTCATCATCATCTTCTTCTTCATCATCATCTTC 457
223 rGlyAsnSerAlaIthrAsnGlyAlaAlaIleCysCylleSerAspLeuA 240
456 TTCTTCTTCATCATCATCTTCTTAATCATCAATTCGTGCTTCATGCCTT 407
240 snThrSerCluLySGlySerLeuSerLeuAlaCysasnGlnIurThLeu 256
406 CATCATCTTCATCATCATCTTCTTCTTCTTCATCATCATCTTCATCAT 357
257 PheaIleSerAsnSerAlaLysGlyAsnGlyAlaAlaIetyAlaIyShI 273
356 TCCTCATCATCTTCTTCATCTTCTTCTTCATCATCA..... 321
273 sMeValLeuArGtyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
320TCTTCATCATCATCTTCTTCTT 299
290 ysIIeGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeu 306
298 CTTCATCATCATCTTCTTCTTCATCATCATCATCTTCTTAATTCAT 249
307 AlaGlyGlyGlySerValLeuPheGlnAsnAsnSerGlnArGThsSerAs 323
248 TGCTGCTTCATCGTCCTTCATCATCTTCATCATCATCTTCTTCTTC 199
323 pGlnGlyLeuValArgAsnAlaAlaIetyrLeuGlnGlyAspAlaIleLeuS 340
198 ATCATCATCATCTTCTTAATCA...TAATCGCTGCTTCATGCTTCTCAT 152
340 eSerLeuGlnAlaArGasnGlyAspIleLeuPhePheAspProIleVal 356
151 CATCT..... 147
357 GlnGlySerSerLeysGlySerProLeuProSerSerLeuGlnAlaSe 373
146 ...TCATCTTCATCATCTTCTTCTTCTTCATCATCTTCTTCATCATCTTC 100
373 rValTrserProThrProAlaThrAlasePrProLeuValIleGlnThrs 390
99 TTCATCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCT 50
390 ePrlAsnaArgseValIlePheSerSerCluArgLeuSer 403
49 CATCATCATCATCTTCTTCATCATCTTCTTCTTCTTCTTC 9

intermodal stem segments"
 BASE COUNT 172 a 179 c 95 g 219 t
 ORIGIN

alignment_scores:
 Quality: 100.00 Length: 200
 Ratio: 0.870 Gaps: 12
 Percent similarity: 57.500 Percent identity: 28.500

alignment block:
 US-09-677-752-4 x AW694070 ..

Align seg 1/1 to: AW694070 from: 1 to: 665

```

359 SerSerSerSerSerSerProleuProleuSerSerSerSerSerSerValTh 375
|||||.....|
7 TCTTCTCATCCCATCACAACCCCTTCAGAAATCGCTTCGCGCATCG 56
|||||.....|
375 rSerProThProAlaThAlaSerProleuValIleGlnThSerAlaA 392
|||||.....|
57 CAACATTCACCAATTCACCTCGCATTCATCATC..... 93
|||||.....|
392 snArgSerValIlePheSer.....SerGlnArgLeuSerGlnGlu 406
|||||.....|
94 .....GTGATTTCTCTCTCTCTGCTTCGCAAAATCATCAACGAAGAA 135
|||||.....|
407 LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGlnLeu 423
|||||.....|
136 CGAACAGCAGCAGCTTCTTCCCTTCATCAAGATCC.....GA 173
|||||.....|
423 sSerGlyArgLeuValLeuLysAspArgAlaVal..... 434
|||||.....|
174 TTCAGAAATATCATCTCAACGACCAATCATCTCTGTTACTTACTT 223
|||||.....|
435 .LeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuIleMet 450
|||||.....|
224 CCATCAAAACCCCGA.....CCCTATTTCTCTCATCTTC 261
|||||.....|
451 GluAlaClyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr 467
|||||.....|
262 TTCACGCTTACTCACTTCAATGACAAATCGAATCAACACCGCAACT 311
|||||.....|
467 uSerIleProLeuHisSerLeuAspThrGluLysSerValThrIleHis 484
|||||.....|
312 TCACAAGAAATTC...TCCATCGTCGCTTCTTCATTCACCAACG 358
|||||.....|
484 lArProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGlu 500
|||||.....|
359 CTAACTCTTCACTCTTCAAAACTCTCTTTTGT.....GAATC 399
|||||.....|
501 AsnPheTyrGluAsnValGluLeuLeuSerLysGlnAsnAsnIlePr 517
|||||.....|
400 GAATTCAGAAATCG.....CAGCTCAGTTCTCTCAATTCGGTGTAA 443
|||||.....|
517 oLeuLeuThrLeuSerLysGlnSerHisLeuHisLeu.....ProA 532
|||||.....|
444 CGCTCTT.....CCCATATTCGCTCTCTGGGCTCTA 475
|||||.....|
532 spGlyAsnLeu.....SerSerHisPheGlyTyrGlnGlyAspThr 546
|||||.....|
476 ATCATGGGTTAGAGATTCATGACATATGAT...CAAGGGGATTTCTCT 522
|||||.....|

```

seq_name: gb_gss25:A2671472

seq_documentation block:
 LOCUS A2671472 883 bp DNA GSS 14-DEC-2000
 DEFINITION ENTJUT77TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, DNA sequence.
 ACCESSION A2671472
 VERSION A2671472.1 GI:11808738
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 883)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000)

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: bjoftus@efr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
 Seq primer: M13-Forward
 Class: shotgun

High quality sequence start: 31
 High quality sequence stop: 867.
 Location/Qualifiers

FEATURES

source

1..883
 /organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOST1; Site:1; Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaundin and B. Bartell, Oxford University Press, 1999)."

BASE COUNT

368 a 122 c 133 g 260 t

ORIGIN

alignment_scores:
 Quality: 100.00 Length: 232
 Ratio: 0.870 Gaps: 7
 Percent similarity: 49.569 Percent identity: 22.414

alignment block:
 US-09-677-752-4 x A2671472/rev ..

Align seg 1/1 to reverse of: A2671472 from: 1 to: 883

```

154 AlaphSerLeuGlnHisAsnTyrLeuPheThrAlaPheGlnGluAsn 170
|||||.....|
541 GCATTGGCTTATACATCTTATATTGGACAGACCTAATATAC.. 494
|||||.....|
170 rSerLysGlyAsnGlyAlaIleGlnAlaGlnThPheSerLeuSera 187
|||||.....|
493 .....TTATTGATTCAGACT 478
|||||.....|
187 rGAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGly 203
|||||.....|
477 CAAATTGGAGTCTAATTGGCAAT.....CCTAATCTTATGAT 440
|||||.....|
204 GlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnValAsnProle 220
|||||.....|
439 ATGCTATATATATGTCCAAT..... 419
|||||.....|
220 uPhePheThrGlyAsnSerAlaThrAsnGlyAlaIleCysCysIles 237
|||||.....|
418 .....GTTGAACATATATG 406
|||||.....|

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L10M502 row: h column: 18
 High quality sequence stop: 605.
 Location/Qualifiers

FEATURES

source

1. 774
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3829697"
 /clone_lib="NIH_MGC_58"
 /tissue_type="hypertrophoma"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgctcgcc); Site_2: SfiI (ggccatagcgc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-ATCTAGAGGCGCGCGGACATG-dt(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.35 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 89 a 297 c 101 g 287 t
 ORIGIN

alignment_scores:

Quality: 101.00 Length: 247
 Ratio: 0.808 Gaps: 13
 Percent Similarity: 50.607 Percent Identity: 28.340

alignment_block:

US-09-677-752-4 x BF030678

Align seg 1/1 to: BF030678 from: 1 to: 774

```

331 iletyleuglulysaspalaialeuserseleuglualaargangl 347
      ::::::::::::::::::::
75  ctcttctctctctctctctctctctctctctctctctctctct 113
      ::::::::::::::::::::
347 yaspilaleupheasproulevalgin.gluuserseleuglu 363
      ::::::::::::::::::::
114 .....ctcttctctctctctctctctctctctctctctctct 156
      ::::::::::::::::::::
364 serpro.....leuproserseleuglualaservalthrserpr 377
      ::::::::::::::::::::
157 tcccttcctctctctctctctctctctctctctctctctctct 206
      ::::::::::::::::::::
377 othproalathralaserproleuvalleglnthseralasnarg 394
      ::::::::::::::::::::
207 ctctctctctctctctctctctctctctctctctctctctct 256
      ::::::::::::::::::::
394 ervalillepserseleugluargleuserglugluylsthrproasp 410
      ::::::::::::::::::::
257 ctctctctctctctctctctctctctctctctctctctctct 306
      ::::::::::::::::::::
411 asnleuthrserleuglnleuglnproileugluuserseleugla 427
      ::::::::::::::::::::
307 tctctctctctctctctctctctctctctctctctctctctct 347
      ::::::::::::::::::::
427 uvalleuglulysaspargalavalleuseralaproserseleugl 444
      ::::::::::::::::::::
348 ctctctctctctctctctctctctctctctctctctctctct 373
      ::::::::::::::::::::

```

seq_name: gp_est89:BF570499

```

444 roglualaleuileuilemetglualaglythrserleuglthrser 460
      ::::::::::::::::::::
374 ctctctctctctctctctctctctctctctctctctctctctct 411
      ::::::::::::::::::::
461 asplleuglualaleuileuileuileuileuileuileuileuile 477
      ::::::::::::::::::::
412 gctgtgtcgcgttcgagcgccgctcctctctctctctctctctc 461
      ::::::::::::::::::::
477 uylsersvalthrllelialapronsluserle.....g 490
      ::::::::::::::::::::
462 tgcactagagggcgcctggcattcattgttactctgtgtttacacca 511
      ::::::::::::::::::::
490 lnyslleupheuserasnerglaspgluasnphetyrgluasnval 506
      ::::::::::::::::::::
512 agaaaatattcggggtcactgctggcagccaccacactcacagaa..atc 558
      ::::::::::::::::::::
507 gluleuser.....lysglulinasnleuileuileuileuileuile 521
      ::::::::::::::::::::
559 tccagtgtagtccccccttaaaaacagcccttcattctctctc..... 600
      ::::::::::::::::::::
521 userlysglulnserleuileuileuileuileuileuileuileuile 532
      ::::::::::::::::::::
601 .....ccattcagccattgttatgttatgccatggcctcaagatcag 643
      ::::::::::::::::::::
533 ....glyasnleuserseleuileuileuileuileuileuileuile 543
      ::::::::::::::::::::
644 agtgtggaac...tcattcagccctggcagatctctgtcagccgagac 690
      ::::::::::::::::::::
544 .....asprthrphesertrlys 550
      ::::::::::::::::::::
691 gggtcttcgagggctcgaacccctttctctcttgagaga 729
      ::::::::::::::::::::

```

seq_documentation_block:

LOCUS BF570499 1466 bp mRNA EST 12-DEC-2000
 DEFINITION 602075417FL NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4242700 5',
 mRNA sequence.
 ACCESSION BF570499
 VERSION BF570499.1 GI:11644211
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 1 (bases 1 to 1466)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DC/DTP
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
 Plate: L10M502 row: a column: 05
 High quality sequence start: 21
 High quality sequence stop: 112.
 Location/Qualifiers

FEATURES

source

1. 1466
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4242700"
 /clone_lib="NIH_MGC_62"
 /tissue_type="melanotic melanoma, high MDR"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggccgctcgcc); Site_2: SfiI (ggccatagcgc);
 Double-stranded cDNA was prepared from cell line RNA. 5'


```

111      |||      |||||:::      ::      |||:::
423 TCTACCACTAGCGATACCAACTGCATCTTCCACTGACCTCACCACCATC 472
415 nleuglnproillegluleulserrglyagluvalleuLysAspa 432
    : |||      :|||
473 AACCCAACCCCAACTCTACTACCAAGGT..... 503
432 rgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
504 ..... CCTCACACTCA..... 515
449 llemetglnlacllythrserleulythrsersaspleuLysleuAl 465
    :|||:::|||||:::|||||:::
516 ..... ACATCTACTACCATGACAGCACCAACCAACCCATCAG 556
465 athLeu... SerleProLeuHisSerLeuAspThglu..... 477
    ||| ||| ||| |||||:::
557 AGACCTTCACCTCCTACATCTACCTACCCAGACACCAAGTTCAACCCCT 606
478 .. LysSerValThrIleHisAlaProAsnLeuSerIleglinsllephe 493
    :|||:::|||||:::|||||:::
607 ACAACACAGTGACTGCCAACTCA.. AACCAACACTACAGGAC..... 647
494 LeuSerAsnSerGlyaspGluasnPheTyrgluAsnValGluleuLeuSe 510
    |||||::: |||::: |||||::: |||||:::
648 ... TCAACTCAACATCTACTACACACAGACAACCAACCCACCTACA 694
510 rlygluglinsnsnliepro.leuLeuthrLeuSerlysglulinSer 526
    :|||:::|||||:::|||||::: |||||:::
695 GGAACAGAGACCTTACCTCCAGATTCTACCACTTACACAGCCCCAAGTTC 744
527 HisLeuHisLeuproAspglyAsnLeuSerSerIsphegy..... 540
    :|||::: |||||:::
745 ATCTTCAC..... CACTCGGGACCAACCC 770
541 . TyrGlnLysTrp..... ThrPheSerTrpLysAsperispsglug 555
    ||| ||| ||| |||||:::
771 CTACACTTAACACTGGTTCACAACACACCCCTACTCCATAAATCTCGG 820
555 lyHisSerLeuIleAlasnTrpThrProLysAsnTYValProHispro 571
    |||||::: |||||::: |||||:::
821 GCCACCCCAGAATACCTAATCGGTTCACAACAACACACCTCAGAAA 870
572 GluATrglinSerThr 576
    ::: |||||:::
871 CACTCAACAACATCA 885
seq_name= gb_gss31:CNS03JGH
```

seq_documentation_block:

LOCUS	1048 bp	DNA	GSS	17-MAY-2000
DEFINITION	Tetraodon nigroviridis genome survey sequence 17 end of clone 03IC18 of library G from Tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL246842			
VERSION	AL246842.1			
KEYWORDS	GI:7967854			
SOURCE	GSS: genome survey sequence.			
ORGANISM	Tetraodon nigroviridis.			
	Tetraodon nigroviridis.			
	Tetraodon nigroviridis.			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinoptera; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha; Holacanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.			
REFERENCE	1 (bases 1 to 1048)			
AUTHORS	Roest-Crolius,H., Jaillon,O., Dasilva,C., Fitzames,C., Fisher,C., Bouneau,L., Billault,A., Quelet,F., Saurin,W., Bernot,A. and Weissbach,J.			
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis			
JOURNAL	unpublished			
REFERENCE	2 (bases 1 to 1048)			
AUTHORS	Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,			

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 SOURCE
 BASE COUNT
 ORIGIN
 alignment_scores:
 Quality:
 Ratio:
 Percent Similarity:
 alignment_block:
 US-09-677-752-4 x CNS03JGH
 Align seg 1/1 to: CNS03JGH from: 1 to: 1048
 134 Glyceratephosphocysarginine synthase, cytoplasmic isoform
 70 GCTTCATGTCGTTCCATCAGAGATTGCGCTCGAGTGAAGTGAAT 119
 150 eserAlaSPAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheG 167
 120 CTTTTCGGAGTGGAGACATCCGATCTTACTCTA..... 156
 167 LUGluAsnSerSerLysGlyAsnGlyAlaAlaGlnAlaGlnThrPhe 183
 157GATGTGACCTGATTGAT..... 174
 184 SerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAs 200
 175GATTCATACCTGTTCCACAGAGAGCAGAGA 206
 200 PLeuAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerLysAsnV 217
 207 GCTGGTCCGAGAGCCTGCTGCGCTTAGTGGC.....TCTGGGACA 247
 217 AlaAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaAlaIle 233
 248 TTTCT.....GATTTGAACCTCCAGT..... 270
 214 CysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSerLeuAl 250
 271GAAGATCTGACGGCGTCACACTTGTGAAGATGGCTCAGACT 311
 250 acyAsn.....GlnGluThrLeuPheA 258
 312 CTTTCACAGGTGACATTCGTAGGTCGGAGATTCACTAGCTACGTCGGT 361
 258 IAsnSerAlaLysGluLysGlyAlaIleTyrAlaLysHisMet 274
 362 CATCTCCGAGAGCAGAGACCTCGGGTGCCTCTTACAGTGGGGA 411
 275 ValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysI 291
 412 TATGATCCGGGTCGGAAATTTCGACATTCATCTCTGATCTGGATCCG 461

```

/db_xref="taxon:4513"
/clone="HV_CEB0017K08f"
/clone.lib="Hordenum vulgare seedling green leaf EST
library HVCNDA0004 (Erysiphe infected & control)"
/tissue.type="seedling green leaf"
/lab.host="rfc121"
/motif=vector: lambda2AP: Site_1: Ecop1: Site_2: Xho1
BASE COUNT      281 a      743 c      35 g      599 t
ORIGIN

```

[illegible]

```
alignment_block:
US-09-677-752-4 x BF064737 ..
```

Align seg 1/1 to: BF064737 from: 1 to: 1688

[illegible]

```

513  GlnAsnAlaIleProLeuLeuThrLeuSerGlyGlnSerHisLeuN1 529
      ::::::::::::::::::::|||  |||  |||
1219  TGTCACTGTATTCACATTTCTATTTCTGSSC.....TTACA 1253
529  sLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspTrpT 546
      |::::::::::::::::::|||  |||
1254  CTAC.....TGTGTCTCANTCTCCACACTGCCATC..... 1284
546  hrPheSerTrpGlyAspSerAspGlnGlyHisSerLeuIleAlaAsnTr 562
1285  .....CACTTGTCCACTCTCT..... 1300
562  rPheProGlyAsnTyrValProHisProGlnArgGlnSerThrLeuValA 579
      |||  |||:::  |||  ::|  |||  |||:::
1301  .ACACATCTTCTCTACAGACACCACTGCACAGCTATACACCTGCACT 1349
579  IaAsnThrLeuTrasnThrTyrSerAspMetGlnAlaValGlnSerMet 595
      ::::::::::::::::::::|||  |||
1350  CATCTTCACTC.....ATCACTGTAC 1372
596  ILeaSnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySe 612
      ::|  ::::::::::|||  |||  |||  ||
1373  TCACTGCCACTCTGTCC.....TTTCCACACTGTCTATC 1407
612  rAlaValSerAsnLeuPheTyr.....A 620
      |::::::::::::::::::|||  |||
1408  TATGTATCTCTCATTCACCTCTATCACTATCACTACATCTTACACACTC 1457
620  IaHisAspSerSerGlyAspProIleAspAsnTrpHisHisArgSerLeu 636
      |||  |||  |||:::  |||  |||  |||
1458  GTCATTCCTCATCAACATCACTCACTTCACTCTCTCTTTCACTACTACTN 1507
637  GlyTyrLeuPheGlyIle.....SerThrHisSerLeuAspAspHis 650
      ::|  ::::::::::|||  |||  |||  ||
1508  CTCTGTGCATCTGCATCACTACCCCTTCTCTCTCTCACTGTCACTGCCA 1557
650  sSerPheCysLeuAlaIleGlnGlyLeuLeuGlyLysSerSerAspSerP 667
      |||  ::::::::::|||  |||:::
1558  TTCTCACTATCACTTACATCTCAT..... 1582
667  heIleThrSerThrGlyThrThrSerTyrIleAlaThrValGlnAlaGln 683
      |||  |||:::  |||  |||:::  |||  ::|
1583  ....ACTGCACCTGTATCCACACACTTACTCTCTCACTCACTATCTCAT 1627
684  leuAlaIleThrSer 687
      |||  |||  |||
1628  CTTCACACACTCN-1639

```

seq_name: gb_est85:BF279812

```
seq_documentation_block:
```

DEFINITION GA__Eb0033P1

ACCESSION	BF279812
VERSION	BF279812 1

KEYWORDS EST. *Coscinium* and *Solirice*

ORGANISM	Gossypium	ar
Eukaryota:		

Magnoliophyta
Malvales; Malvaceae

REFERENCE 1 (bases 1)
AUTHORS Wing, R.A., 1

TITLE	An integrat
-------	-------------

JOURNAL Unpublished

Clemson Univ

100 Jordan I
Tel. 864 650

Tel: 864 656 7288

17-NOV-2000
Library Gossypium

ta; Spermatophyta;
idae; eurosids II;

Simmons, J., Henry
nt, and evolution

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288

531 oaspglyasnleuserSerHisPheGlyTyr 541
 174 C.....CTTCTCTCCATCATCATCTCTAT 150
 seq_name: qb_gss17:A2218282

seq_documentation_block:

LOCUS A2218282 576 bp DNA GSS 09-JUN-2000
 DEFINITION Sheared DNA-81G10.TR Sheared DNA Trypanosoma brucei genomic clone
 Sheared DNA-81G10, DNA sequence.

ACCESSION A2218282
 VERSION A2218282.1 GI:8436082

KEYWORDS GSS.
 Trypanosoma brucei.

SOURCE Trypanosoma brucei
 ORGANISM Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE

1 (bases 1 to 576)
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Sub, E., Malek, J., Fujii, C.,
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
 Fraser, C. and Adams, M.

TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library

JOURNAL Unpublished (1999)
 COMMENT Other_GSSs: Sheared DNA-81G10.TF

Contact: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for
 distribution through Research Genetics, Alabama, USA. Sheared DNA
 end sequences search page: <http://www.tigr.org/cdb/mdb/cddb/>.
 Seq primer: M13-Reverse
 Class: shotgun.

FEATURES

SOURCE

Location/Qualifiers

1..576
 /organism="Trypanosoma brucei"
 /strain="TREU927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-81G10"
 /clone_11b="Sheared DNA"
 /note="Vector: pUC18; Site.1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + 1 method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaundin and B. Barrell, Oxford University
 Press, 1999)."

BASE COUNT 119 a 132 c 121 g 204 t
 ORIGIN

alignment_scores:

Quality: 106.50 Length: 224
 Ratio: 0.772 Gaps: 11
 Percent Similarity: 61.607 Percent Identity: 21.875

alignment_block:

US-09-677-752-4 x A2218282 ..

Align seg 1/1 to: A2218282 from: 1 to: 576

285 IlaasnasuseralalsileglyAlaAlaAlaAlaInserglyc1 301
 6 TTCCATTCATCATCTTCATTCAGTACCCTGTGATGTTCATCATCATCGGT 55

301 yserleuserlileuAlaGlyGlyGlySerValleuphegnasnas 318
 56 TTCTTGCAGTCACTGTGATGTTCATCATCTGCG 90
 318 erglnArgThrSeraspGinglyLeuValArgasnaAlaIleTyrleuglu 334
 91 TTTCATTGACGATCACCCTGTG...ATGTTCCAAATCATCGGTTTCATTGACG 137
 335 lysAspAlaIleleuserSerleugAlaAlaArgasnglyAspIleleugh 351
 138 TCACCTGTGATGTTCGAATCATCAGTTTCATTCACAGCTGTGATGTCGAATCA 185
 351 ephasproIleValGlnGlyuserSerlysguserProleupros 368
 186TTGCATCATCGGTTTCATTGACGATCACCCTGTG... 218
 368 erSerleuglnAlaSerValThrSerProThrProAlaThrAla..... 382
 219 ..ATGTTGCAATCATCAGTTTCATTCACAGCTGTGATGTCGAATCA 266
 383SerProleuValIleGlnThrSerAlaasn...Ar 393
 267 TCGGTTTCATTCAGTCACTGTGATGTTCGAATCATCAGTTTCATTGCA 316
 393 gSerValIlephasSerSerGlnArgleuserGinglyulysThrProA 410
 317 GTCACTGTGATGTTCGATCATCATCGGTTCA..... 347
 410 sPasnleuThrSerGlnleuglnProIleGlyleuysSerGlyArg 426
 348TTGCAGTCACCTGTGATGTTCGAATCATCA... 377
 427 leuValleuLysaspArgAlaValleuserAlaProSerleuserGlnas 443
 378GTTTCATTGACAGTCACCTGTGATGTTCGAATCA... 409
 443 pPro.....GlnAlaLeuLeuIleMetGlnAlaGlyThrSerLeuL 457
 410 ATCGGTTTCATTCAGTCACTGTGATGTTCGAATCATCGGTTTCATTGCG 459
 457 ystThrSerSerAspIleuLysleuAlaThrleuserIleProleuHisSer 473
 460 ACTGACCTGTGATGTACAA.....TCATCGGTTTCATTCAGTCA 500
 474 ..LeuAspThrGlnLysSerValThrIleHisAlaProasleuserI 489
 501 CCTGTGATGTTCGAATCATCGGTTTCATTCAGTCACTGTGATGTGCGCA 550
 489 eglnLysIlePheleuserasn 496
 551 ATCATCGGTTTCATTGACGATCA 572
 seq_name: qb_est93:BF864174

seq_documentation_block:

LOCUS BF864174 1166 bp mRNA EST 19-JAN-2001
 DEFINITION 963046G03.Y1 C. reinhardtii CC-1690, stress condition I, normalized
 , lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF864174
 VERSION BF864174.1 GI:12254318

KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 1166)
 Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
 Lefebvre, P., McDermott, J.P., Shiger, J., Sillflow, C. and Stern, D.

TITLE

Analyses of the Chlamydomonas reinhardtii Genome: A Model,
 Unicellular System for Analyzing Gene Function and Regulation in
 Vascular Plants, project phase 3
 JOURNAL Unpublished (2000)

```

410 ASPAspLeuThrsGlnLeuGlnProIleGluLeuLysSerGly 426
    :|||:|||||
628 GTTCTTGTCTCT.....TCTGTAC 650
426 GluValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGln 443
    :|||:|||||
651 ATTGTG.....GATGTAATGATGCTCTTGTCCACATGTT 668
443 sPrroGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuLysThr 459
    :|||:|||||
689 CACCA.....TTCTGCCCATCTTCACACACACAGCTCTCTCTCC 729
460 SerAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuSph 476
    :|||:|||||
730 TCTTCTCACTCTTCTCT.....TCTTCCAAACATCTTCGGCTGTAC 776
476 r...GluLysSerValThrIleHisAlaPro.....AsnLeuSerIleG 490
    :|||:|||||
777 ATCAGAAAGATCCGCTCTCTCTCTCCACGATCATCTCTGTCTTC 826
490 InLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
    :|||:|||||
827 CTCTCTACATTTGCTCTCTTCGATGCTTCCTCTTC 864
seq_name: gb_gss23:A2545202

```

```

seq_documentation_block:
LOCUS      A2545202      874 bp      DNA      GSS      14-NOV-2000
DEFINITION ENTDR55TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION  A2545202
VERSION    A2545202.1 GI:1165757
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 874)
AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
TITLE     Determination of clone end sequences from Entamoeba histolytica
          HM1:IMSS sheared DNA library
          Unpublished (2000)
JOURNAL    Contact: Brendan J Loftus
COMMENT    Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: b.loftus@tigr.org
          Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
          DNA library
          Seq primer: M13-Forward
          Class: Shotgun
          High quality sequence start: 17
          High quality sequence stop: 831.
          Location/Qualifiers
            1..874
              /organism="Entamoeba histolytica"
              /strain="HM1:IMSS"
              /db_xref="taxon:5759"
              /clone_lib="Entamoeba histolytica Sheared DNA"
              /note="Vector: pROSI; Site 1: Bst I; Constructed at The
              Institute for Genomic Research (TIGR), Rockville, MD.
              Genomic DNA isolated from broth cultures of E. histolytica
              using a method described by Clark and Diamond (Clark,
              C.C., and Diamond, L.S. (1993) Entamoeba histolytica: a
              method for isolate identification. Exp. Parasitol.
              77:450.). The DNA was mechanically sheared to give a
              tight size distribution (~2 kb). The v + i method used for
              the library construction is described in detail in Smith,
              H.O. and Venter, J.C. (Making small insert libraries for
              whole genome shotgun sequencing projects. In Genome
              Sequencing: A Practical Approach, eds. M. Vaubin and B.
              Barrell, Oxford University Press, 1999)."

```

```

BASE COUNT      332 a      99 c      136 g      307 t
ORIGIN
alignment_scores:
  Quality: 112.00      Length: 392
  Ratio: 0.636      Gaps: 17
  Percent Similarity: 44.898      Percent Identity: 22.194

```

```

alignment_block:
US-09-677-752-4 x A2545202/rev

```

```

Align seg 1/1 to reverse of: A2545202 from: 1 to: 874

```

```

155 PheSerLeuGlnHisAsnThrLeuPheThrAlaPheGluGluAsnSerSe 171
    :|||:|||||
854 TTATATATAAACAATTATATTA.....GAACATTATTC 820
171 rLysGlyAsnGlyAlaIleGlnAlaGlnThrPheSerLeuSerArg 188
    :|||:|||||
819 TAATCTTAACCAAAACCAATTTTATTCTTCTCCATTAATTTTAACA 770
188 snValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGly 204
    :|||:|||||
769 ATCTTCTCTTATACATTTGCTCATCTTACATCT.....CTTAT 732
205 AlaIleCysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeu 221
    :|||:|||||
731 .....TGCTTGAAATTAAGTCTCTCT.....CTTAT 703
221 ePheThrGlyAsnSerAlaThrAsnGlyAlaIleCysCysIleSer 238
    :|||:|||||
702 TTATGATTCAAAT.....GTAATAAACTTTAGTG 674
238 sP.....LeuAsnThrSerGly 245
    :|||:|||||
673 ATTATTCATTAGTGGATTGGAACATTTGTAAGATTGCTCTTAAGC 624
246 SerLeuSerLeuAla.....CysAsnGlnGluThrLeuPheAlaSer 260
    :|||:|||||
623 TCAATTTAATACAGCTCTCGAATCT.....600
260 nSerAlaLysGlyLysGlyAlaIleThrAlaLysHisMetValLeu 277
    :|||:|||||
599 .....GAAATGGATCGAATTAATATATCCACAAATTTATATATT 560
277 rGTYAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGly 293
    :|||:|||||
559 CATAT...TGCTTGTAGTGTGTCGATGCTTATTAATATACCAAC 513
294 AlaIleAlaIleGlnSerClyLysSerLeuSerIleLeuAlaGlyGlu 310
    :|||:|||||
512 ACTATTTCAAGAGTAAGCTCTTAGTATCAAGTATTAAGCA..... 468
310 ySerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeu 327
    :|||:|||||
467 .....AAATTAATTA 458
327 aArgAsnAlaIleThrLeuGluLysAspAlaIleLeuSerSerLeuGlu 343
    :|||:|||||
457 TCAATGGGATCTTCATTAAGTATTAATTAATTAATTTCTCATTAACA 408
344 AlaArgAsnGlyAspIleLeuPheAspProIleValGlnGluSerSe 360
    :|||:|||||
407 ATTACAGTCATCTACTGTAATTAATTAAGGCACACACTGCACTCAAT 358
360 rSerLysGluSerProLeuProSerSerLeuGlnAlaSerValThrSer 377
    :|||:|||||
357 AAATGCTATTAATCATCATCTGATCTATTT.....C 326
377 rThrProAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArg 393
    :|||:|||||
325 CAATTTAATCTTCATCACTACATCAATTAATTAATGAGCGTTAATAA 276

```

212 TCACAGTGGAGAGCTTCTTGACCAACTGCTTGTGCCCCCTGCACACAT 261
383 erProleuValIleGInThrSerAlaAsnArgSerValIlePheSer 399
262 CTCCTATTATGCTACCAACCATGACCAACAGTATTATGACCATGCA 311
400 GlnArgLeuSerGlnGluGlyThrProAspAsnLeuThrSerGln 416
312 CACCAACAACAACAACAACAATGCTCTT.....CAAAATAT 349
416 uGInGInProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg 433
350 TCACACACCA..... 359
433 laValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuIle 449
360CGACGACCTTCT.....CAACCAACAATATATACAGT 392
450 MetGlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 466
393 CCAAGTGTAGTATCATCTGCACGACATCATCT.....GCAAC 430
466 rLeuSerIleProLeuHisSerLeuAspThrGluLysSerValThrIle 483
431 ATCTACTCAACCA.....C 444
483 laAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 499
445 AATGCCACAATCATATGCGCAGTATTACTACTACTAATACCAATGAC 494
500GluAsnPhenylGluAsnValGlu.....LeuLeu 510
495 AATAAAGAAAGAAAGAACTTCTGAGAGAAAGAAAGAAAGAACTTCTGAGAA 544
510 rLysGlnGlnAsnAsnIle 516
545 AAGAGACAAACGATATT 563
seq_name: gb_gss32:CNS05015
seq_documentation_block:
LOCUS CNS05015 1086 bp DNA GSS 26-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
03A114 of library B from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL348697.1 GI:8242467
VERSION AL348697.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Auripeterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Perciformes;
Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 1086)
Roeest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 1086)
Roeest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1086)
Genoscope.
TITLE JOURNAL
AUTHORS
REFERENCE
JOURNAL
TITLE
Direct Submission

JOURNAL Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
FEATURES
source
1..1086
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="03A114"
/clone_id="B"
/note="Genoscope sequence ID : C0NB034BE07C1-end : T7"
BASE COUNT 311 a 241 c 239 g 282 t 13 others
ORIGIN
alignment_scores:
Quality: 116.00 Length: 300
Ratio: 0.758 Gaps: 17
Percent Similarity: 51.000 Percent Identity: 25.000
alignment_block:
US-09-677-752-4 x CNS05015 ..
Align seg 1/1 to: CNS05015 from: 1 to: 1086
384 ProleuValIleGInThrSerAlaAsnArgSerValIlePheSerGln 400
105 CCATATTATTAAANCAAGTCAATGACACAAACGATGTTGACAGCCC 154
400 uArgLeuSerGlnGlu.....GluLys.ThrProasp 410
155 TTTTGTCTATTTCAGACTTCTGCGCGCGCGCTGTGAAGAACCATGGCG 204
411 AsnLeuThrSerGlnLeuGlnPro...IleGluLeuLys...SerG 425
205 CGGTTGGATGCGCTGCCAAGAAAGAAAGTTATTGAGCTCGGGAATCTGG 254
425 yArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerG 442
255 TTGAGATTTCCTAAATCAACACAGCTCCTCGAGCTCGAGACATTAAAG 304
442 lnsAspProGlnAlaLeuLeuIle..... 449
305 TGTCTCGCAGCGCAATCTACTCTGCTGAGAGGATTTCATGAGACCA 354
450MetGlnAlaGlyThrSerLeuLysThrSer 460
355 CACGAGAGAGTACGCCCTGTGAGGCTGGAAGC.....ACCTCGTC 395
460 rAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrG 477
396 ATCAGAGCAGATGCAM.....CCTCAGTCGGAACATCAGCAG 433
477 lLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
434 GAAGAGAGAGCAATATTTCACGTTCCAAACATTGCCAGCAACTTCTCAT 483
494 LeuSerAsn...SerGlyAspGluAsnPhenylGluAsnValGluLeu 509
484 CATGTAACCTTCTTGAACAGGAAATTTACAGACACGCTTACATT 533
509 uSerLysGlnGlnAsnIleProLeuLysThrLeuSerLysGlnGln 526
534 TTCGAGCGCCGATGCAAAATTTGCC.....ACAAAGGAAGCA 571
526 eHisLeu...HisLeuProAspGlyAsnLeuSerSerIlePheGlyTr 541
572 GCGAGCTGGAACCAACATTAATGAATAAAGAGAAAA...GGGTAT 617
542 GlnGlyAspTrp..... 545
618 TCAGATTGTGGCGTCACCTCTTTCGACAGAACACCAACAATAATATC 667

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81601
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,947 <REA>
A:Cross-references: GB:AE002182; GB:AE002161; NID:q7189140; PIDN:AR38083.1; PID:q718914
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_19; CP0213

Query Match 11.5%; Score 586; DB 2; Length 947;
Best Local Similarity 25.2%; Pred. No. 3,5e-28;
Matches 250; Conservative 173; Mismatches 390; Indels 178; Gaps 47;

14 LSGLAAREVPSRIELMPNSVDPPTKESLSNKI---SLTGDTHNLTCYLDNLRIILAILOK 70
Db LSGIHSGEDLEFLTRSSSPTKTITSLRKDFIVCDPAGNSIIRPGAAFLNLGDLFFINS 89
Qy 71 TPNEGAAYITDYLSPFDQKRGIVFAKULTPESGALIGYASPNPTVE-----IRPT 123
Db 90 TPL--AALTFKNI--HLGARGAGLSESNT-----FKGLHSLVLENNESWGCVLT 137
Qy 124 IGPVIFENNT--CCRPETSSNPNAVANKIREGAIHAQN---LYINHHVDVGFMKNF 176
Db 138 SCDLSFINNTSVLCQNNISYCP-----GGALLQCRKSKALFRRDNGFJLFLEKMK 188
Qy 177 SYVR-----GGAISTANTFVYSENOGFLMDN--TCIQNTAGKGAITAGT--SNF 226
Db 189 AVNODESHPGYGAVSSISP-----GSPITPADNOELFQENEGELGAIYINDOGAITE 242
Qy 227 ESNNDLDFINNACGAGIFSPICLTGNRGNIYVNNRCRKNVTASSEASDGAIV 286
Db 243 ENNFOTTFSSKASFGAVSYRCULYSQKDTLF-----TKNAAKVGAIR-- 291
Qy 287 TRLDVTGNRGRIFFSDNTTKNGAIVAPVYLVLD--NGPTVFINNIA--NNKGAIYIDG 344
Db 292 ADVVHTRDCKGSIIVEEN--SATAAGAIAVNAVDINAQGVPRFINNSALGLNAGAIYMOA 350
Qy 345 T--SNSKISADRAITFENENIVTV--TNANGTSTANPRRNAIIVYASSGELLGAGSS 401
Db 351 TGSILRLHANOQIEFCGKRVRSQFHSINSTSNTFN---NAITIGAPREFSLANEG 406
Qy 402 ONLIFDPI-----EVSAGVSVSEFKADDTGVSFGATVNSADRHORLQKTP--A 454
Db 407 HRICTPDITISATENTNS--LYINHQRLEAGCAVIFSGARLSE--HKENKKTSTINO 463
Qy 455 PLTISNGFLCIEDHQAOLTVNRFTOTGGVVSILGAVLSCYKNGAGNSASNAITLKHIGL 514
Db 464 PVALGCGVLSIEGAILAARFYGEGGLLALGSGSKLTT---QGNKSEKDKIVITNLGF 519
Qy 515 NUSILKSG--AEPLPLMVEPTNNSNNTADTATFSLSDVKL-----SLID--DVGNP 565
Db 520 NLENDSSDPAET-----RATEKASIEISGVPRVYGHTESEYENHEKASKP 565
Qy 566 YESTDTHALSSQPMLSISEASDNLRS---DEMDFSGLVPHYGOGGLTWMATQDP 622
Db 566 YTTISIT---LSAKKLVTAAPRKQDIONLLIASEYWG-----YVQGSSEFMSNDPK 617
Qy 623 EPASSAATIDPOKARNFHRTLLTLWLPAG--YVSPKHSRSLPLANTIGMNL-----LA 674
Db 618 E-----KTIIASMTPTGEFSLDPKRGSGFIPTLIMSTFGSLINISIV 661
Qy 675 TESLKASAEITPDHPWMTGGGLGMVYQDRENNHPCGHMSSSGISACMIAGQHTFS 724
Db 662 NNNYLNNSSEVPLQH--LCVFGGPVYOIMQENKROSSNNLLVOHAGHNVG--ARIPEFS 717
Qy 735 LKRSQYTKRNEYAKNNVSKNYSOGEMLF---SLQEGFLTLKVLGLSYGDHN---C 788
Db 718 TTIISALITOLFSSSSQONVADKSHA---QILIGTVSLNKWQALSLSRSSSTYEDSDYMK 774
Qy 789 HNFYTOGENTLSQGTFRSQTMGAVFFDLPMKDFGTHILTAFLGALGIYSSL--SHT 846

Db 775 HVFPYK---TSRGSWRNYGWSGYSGMGVAY--PKQIRFLKMTPFVDDO--YTKLVQNEFV 828
Qy 847 EVGAYPRSESTKPLINLVLPVIGVGSFNNATOPROAM--TWELAYQPLVROEGIAIOL 905
Db 829 ETGIDPRFFSS--SEKTNLSLPIGIALEMRFFIGSRSLFLOYSTSIKDLRVNPOSSASL 887
Qy 906 LASKGIMFGSGSPSSRHAMS-----YKI 928
Db 888 VLNHYTWDIQCPVPLGKEALNITLNTSTIKYKI 918

RESULT 21

B72077

polymorphic membrane protein G family CP0306 (imported) - Chlamydia pneumoniae (s
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: B72077; B81592
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606

A:Accession: B72077
A:Molecule type: DNA

A:Residues: 1-928 <RNA>
A:Cross-references: GB:AE001628; GB:AE001363; NID:q4376730; PIDN:AD18591.1; PID:q437

A:Experimental source: strain CWL029
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.: Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255

A:Accession: B81592
A:Molecule type: DNA

A:Residues: 1-928 <REA>
A:Cross-references: GB:AE002192; GB:AE002161; NID:q7189226; PIDN:AR38163.1; PID:q718

A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_9; CP0306

Query Match 11.5%; Score 585.5; DB 2; Length 928;
Best Local Similarity 24.9%; Pred. No. 3.7e-28;
Matches 266; Conservative 139; Mismatches 380; Indels 285; Gaps 46;

1 MKKAFEPFLIIGNSL-----SGLAREV-----PSRIPLMNSVDDP--TKESLSNKIS 45
Db 1 MKSSLMHFLISSSLALPLSLNSAFAPAAVEINLGPINSSGFGITPRPAQTNNADGTYIN 60

Qy 46 LTGDTNHLTN-----CYLD--NURYI---LAILQKTPNEGAAYITD----- 82
Db 61 LTGADV--SITNAGSPALATFASCFFKETTGNLSFOHGQYQFLLOINDAGANCTNTAANKLL 119

Qy 83 -----YLSFDTQKRGITFAKULTPESGALIGYASPNPTVEIRDTQPVFENNTCR 136
Db 120 SFEGSYSLDIQT-----TNATGTGA-----IKSTGACSIOGNSYSCF 158

Qy 137 PFTSNPNAAVNKIREGAIHAQNLVNNHHVDVGEKNSYVRGCAISTANTFVSENO 196
Db 159 GONFSNDN-----GALOGSSISLSLNPNT--FAKKAKAQKGAALSTGIGITINNLT 209

Qy 197 SCFLFMNDNICIQNTAGKGAITAGTSNFSNCDLFINNACCA-----GGAISFPIS 252
Db 210 NSASFSEN-----TANNNGAIYTEAS--SFISSNKAISFINNSVTATSAITGCAIY---CS 260

Qy 253 LTGNRGNIYFNNRCRKNVETASSEASDGAIVKTRLDVYTGRRGIFPSDNTKRYGA 312
Db 261 STS-----ADRPVLTLS-----DNGELNLTGNTAITSGGA 290

Qy 313 IYAPVTVLNDGPTVFINNIAANK-----GGAIVIDGTSNKSISADHAAIFENIVTNT 368
Db 291 IYDNLVLSGGFTLFRKNSAIDTAAPLGAIAIADSGSLSLALOGDITFEGNTVYKKA 350

Db 948 VRGSMYAS-----LTGNIIEYVG 964

RESULT 14

A:1591
polymorphic membrane protein G family CP0307 (imported) - Chlamydia pneumoniae (stra
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
A:Accession: A81591
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REV>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AF38164.1; PID:g718923
C:Genetics:
A:Experimental source: strain AR39, HL cells
A:Gene: CP0307

Query Match 12.2%; Score 620; DB 2; Length 930;
Best Local Similarity 27.1%; Pred. No. 2.8e-30;
Matches 252; Conservative 139; Mismatches 374; Indels 164; Gaps 40;

47 TGDTHNLNCTYLDNRLYLALQKTPNEGAAT---ITDYLSEFDTOKEGIFYAKNLTPE 103
Db 50 TADA-NGTNYVLSCNVI-----NDACKGTALTGCCFTETGDLTFGKGYSFSFN-TVD 102
Qy 104 SGGAIGVSPSPYIEIRDT-----IGPIYFENNTCCRFSTSSNPAANVKIREGAIHAQ 159
Db 103 AGSNGAALSTADKALFTGFSNLSFIAAPCTTVAASGKSTLSAGALINTDNCITLFSQ 162
160 NLYINHNDVVGFMKFSYVGAISTANTFVSENOCEFPMONICQIOTNTAGK-GGAI 218
Db 163 NVSNENANN-----GGAI-TAKTILISGNTS-----SITFTSNAKKLGAI 203
Qy 219 YAGTNSFESNNDLFFITN-ACCAGAI-FSPICSLTGNRCNIVFYNNRCKNVTASS 276
Db 204 YSSAASISGNTGQLVPMNNKGTGALGFEEASSITON-SLFFSGN-----TATD 255
Qy 277 EASDGAIV-----TTRLDVTGNRGRIFESDNTITKNNGAIYAPVTVLVNCGPTTYINN 331
Db 256 AAGKGAIVCEKTEGTPTLTISGNKS-LTFAENSSVTGGALCAHGLDLSAAGPTLFSNN 314
Qy 332 IANN-----KGAIVYDGTNSKISADRAHIFENENIVNTNANGTSTANPPRRNATV 387
Db 315 RCGNTAAKGAIALADSGSLSLANOCIDITFLGNTL-----TTSAPSTRNATYL 366
Qy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAGVS-----VSFNKEADQYGSVVFSGATVNS 440
Db 367 GSSAKITMLRAAOCQSIYFYDPIASNTTGASDVLITINOPDSNPLDYSGTIVFSGEKL-S 425
Qy 441 ADFHNR--NLQTKPAPLTLNGLFCIEDHAOLYNNRTOTGGVSLONGAVLSCTYKGA 498
Db 426 ADEAAADNFTSILKOPALASGLTALKGNVELDVNGFTQEGSTLL-----MQGT 477
Qy 499 GNSASNASITTKHIGLNSLILKSGAEIPLLVPEPTNNSNNVTADTAATFSLSDVKLSLI 558
Db 478 KIKADTEAISLTKLVVDLSAL-----EGNKSVSITAGANKTITLTS-PLVFO 524
Qy 559 DDCNSPSTEDTLHALSSQPLSLISEASDQLRSDMDVDFSGLVN-----PHYGWL 611
Db 525 DSSGNF-YESHTINQAF--TQPLVFTAAAT---AASDIYIALTLSPVOTPEPHGYGCH 578
Qy 612 WTGCAKATODEPPASSATITDPOKRNRRHRLILLTLWLPAGVYPSKHKHSPLIANTLKNM 671
Db 579 WEATWA---DTSTAKSGT-----MTWVTGYNPDPERRASVAPDLSMASF 620

Qy 672 LLATESLKNSAELTPSDHPFW---GITGGLGMMVYODPRENHGPFHMRSSGYSAGMIAG 728
Db 621 ---IDIRTLQOIMTSQANSITVOQRGLANSGTANFPHKKSSTNOAFRKSIGTYVGSAAE 677

Qy 729 --QHTFSLKFSQTYTKLINERYAKNNVSSKNVSCO-----GEMFSLQ 769
Db 678 DFESENFVACQLFGKDKDLFIYEN-TSHNVYLSLVLQHRAPFLGGLPMPSPGSIITDMLK 736

Qy 770 EGFLLTKLVGLYSYCDHCHHFFYT-----QGENLTSQSTFRSQTMGCAVFFDLPMK-PFG 823
Db 737 DIPLLNQSLSYTKNDMDTRTYSYPEAQSMTNNSGALD--LGGSLALYLPEKAPFF 793

Qy 824 STHILTAPFLGALGYSLSHFTVEGAVPRFSFRTPLINVLVPIGVKGSFMMATOPQA 883
Db 794 QGYF---PFLFQAVYSNQONFKESGACARAFD-DGDLVKNCSIFVGIRLEKISEDEK-NN 848

Qy 884 WTVELAYQPVLYROEPGIATQLLASKGIW 912
Db 849 FEISLAVIGDYVYRKNRPSRTSLMVSAGSM 877

RESULT 15

D72078
polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
A:Accession: D72078
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lamell, C.; Fan, J.; Ollinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72078
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <ARN>
A:Cross-references: GB:AE001627; GB:AE001363; NID:q4376721; PIDN:AD18590.1; PID:q437
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_8

Query Match 12.1%; Score 617; DB 2; Length 930;
Best Local Similarity 27.0%; Pred. No. 4.3e-30;
Matches 251; Conservative 140; Mismatches 374; Indels 164; Gaps 40;

47 TGDTHNLNCTYLDNRLYLALQKTPNEGAAT---ITDYLSEFDTOKEGIFYAKNLTPE 103
Db 50 TADA-NGTNYVLSCNVI-----NDACKGTALTGCCFTETGDLTFGKGYSFSFN-TVD 102
Qy 104 SGGAIGVSPSPYIEIRDT-----IGPIYFENNTCCRFSTSSNPAANVKIREGAIHAQ 159
Db 103 AGSNGAALSTADKALFTGFSNLSFIAAPCTTVAASGKSTLSAGALINTDNCITLFSQ 162
Qy 160 NLYINHNDVVGFMKFSYVGAISTANTFVSENOCEFPMONICQIOTNTAGK-GGAI 218
Db 163 NVSNENANN-----GGAITT-KTILISGNTS-----SITFTSNAKKLGAI 203
Qy 219 YAGTNSFESNNDLFFITN-ACCAGAI-FSPICSLTGNRCNIVFYNNRCKNVTASS 276
Db 204 YSSAASISGNTGQLVPMNNKGTGALGFEEASSITON-SLFFSGN-----TATD 255
Qy 277 EASDGAIV-----TTRLDVTGNRGRIFESDNTITKNNGAIYAPVTVLVNCGPTTYINN 331
Db 256 AAGKGAIVCEKTEGTPTLTISGNKS-LTFAENSSVTGGALCAHGLDLSAAGPTLFSNN 314
Qy 332 IANN-----KGAIVYDGTNSKISADRAHIFENENIVNTNANGTSTANPPRRNATV 387
Db 315 RCGNTAAKGAIALADSGSLSLANOCIDITFLGNTL-----TTSAPSTRNATYL 366
Qy 388 ASSSGEILLGAGSSONLIFYDPIEVSNAGVS-----VSFNKEADQYGSVVFSGATVNS 440
Db 367 GSSAKITMLRAAOCQSIYFYDPIASNTTGASDVLITINOPDSNPLDYSGTIVFSGEKL-S 425

[illegible]

polymorphic membrane protein G family CP0308 [imported] - Chlamydophila pneumoniae (s

RESULT 12
B81591
C:Species: Chlamydomonada pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 Sequence Revision 31-Mar-2000 #text-change 11-May-2000
R:Accession: B81591
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonyak, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID: 20150255
A:Accession: B81591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-936 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; NID:g7189234; PIDN:AAF8165.1; PID:g718
A:Experimental source: strain AR39, HL cells
A:Genetics:
A:Gene: CP0308

Query Match 12.7% Score 644; DB 2: Length 936;
Best Local Similarity 27.0%; Pred. No. 9,4e-32;
Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

Qy 75 GAATTTDVA,SPFDROKESIVFAKNLTPESGAICY-----ASGNSPTVEIR 121
| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 54 GTTSLSDVSFQNAGALICPLASGCFLGEGDLTFQGQHAKLFATINAGSSACTVAST 113

Qy 122 DTIGPIYIFENN-----CCCRFTSSNMNAVNKRIBEGCAIHAONLYLNHHDDVGFMK 174
| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 114 SAADKKLLFNDSRLSIISCPILLSPTQCQAKSV-----GLSLITGNSQLI-FTQ 164

Qy 175 NFSYVNGAIISTANFVSEENOSCELFMDNICIQNTTACKGAIIYA-GTSNFSESNNCDL 233
| : : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 165 NFESDNGVIINTKN-FLLSGTSGOFASFSRN--QAFTKGGGVVATGTITENSPIGVIS 220

Qy 234 FYTNNAACCAAGALIFS-PICSLTGNRGNIIFYNNRCKNVETASSSESDOGAKIKVT--R 289

Db 652 WQSPHNLFLATLRYOTQOQIAPT-----ASGEATRLFVHONSNDANGFHEATGSLG 705
Oy 725 MIA---GGTFTSLKFSOTYTKLNERAKNNYSKNSCGEMLFS-LOEGFLTLTKVLY 781
Db 706 TTSNTASHNSHSGVNSOLFNSLIESHSDNSVASHHTTALQIINNWLDERFSTSLA-X 764
Oy 782 SYCDHCHHFFYTQGEN--LTSOGTFRSOTMGCAVEFDDLPKPFSGTHILTAFLCALGTY 839
Db 765 SYSN---HHIKASGVSGKIQTEGKCYSTTLGALSCSLQ- WSRPLHFFPFIQAIAYR 820
Oy 840 SLSLHTEVGAIPRFSFTKPLINLVPIGVKSGFMNATQORQATVETLAPVLYRQEP 899
Db 821 SNOTAFQESGDARFESYHKKPLYNLTPGLQIOWAMESKFRPLTYNNIELAYQVLYQONP 880
Oy 900 GATOLLSKGIWFGSGSPSSRHAMSRYKISQOTQPLSMULTLFOYHGFYSSSTFCNYLNG 959
Db 881 EINVLESSESGSWLLSGTTLARNNAIFKGRNOIFLPKLSVLDYOGVSSSTTHYLHA 940
Oy 960 EIALRF 965
Db 941 GTTFKF 946
RESULT 8
F71460
probable outer membrane protein F - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: F71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis. Reference number: A71570; MUID:99000809
A:Accession: F71460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1034 <ARN>
A:Cross-references: GB:AE001360; GB:AE001273; NID:93329342; PIDN:AAC68468.1; PID:9332934
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: ompF

Query Match 17.0%; Score 866.5; DB 2; Length 1034;
Best Local Similarity 28.1%; Pred. No. 2.3e-45;
Matches 293; Conservative 149; Mismatches 406; Indels 193; Gaps 36;

Oy 55 NCYLONLRYILAILLOKTPREGAAVTTIDYLSFFDTQKEGIFYFAKNLTPESGAI----- 108
Db 57 NIVLSMLQ-----SNGTGACTISG-----NTOTQIFSNSTNTTADSGAFDMVTTTS 102
Oy 109 -----GYASPSPTVEIRDITIGPVIEFNNTCCRPSTSSNPNAV-----NK 149
Db 103 PTASDANALLFCNNCTYHNKGGAIIRSG-GPIRFLNNQDVLFPYNNISACAKYVGTGDHNE 161
Oy 150 IREGAIIHONLYINHHNDVYGFMKNFYVYRGGAISTANTFVVSSENOGCLFMDNICI-- 207
Db 162 KNNKGALLVATITLIGNR-TLAFINMMSGCGAISADQISTIDIVKGLFENHNLNH 220
Oy 208 -----QNTAGKGAITYAGTS-NSFESNCDLFFINNACCGAIFSPICSLTNRCNIYF 262
Db 221 IPTYQENNAKGAICSRDLCSISNNSGPIYENYNOGKGCAISATRCVIDNNKERIIF 280
Oy 263 YNRRCKKWTETASSEADGAIVTTRLDVTRGRRIFFSDNTKNYGAIVAPVTVLVD 322
Db 281 SNNSSLG--WSOSSSASNGCAIOTTOGFTLRNNKGSIFPDSNATNAGCAINCICYDIRD 338
Oy 323 NGPTTFINNIANKGAIYIDGTSNKSISADRAH-----IIFENIVTNTVANG 372
Db 339 NGPVYFLNNSA--AMGAAF-----NLKRPASATNYIHTGTDIYVNNNVFTLQNL 389
Oy 373 TTSANPRRNATTVASSSEIILGAGSSONLIYDPIE-----VSNAGV-----V 419

Db 390 GKRKLPHINNHETPTTIS-----LGAKKDTRIYFYDLFQWERVKENTSNNPSPSTRN 445
Oy 420 SENKEADQTSVEFSGATVNSADFHQRNLQTK-----TPAEILTSNEFLCIEDHAQTLVN 474
Db 446 TVNPETESGAVNVS---YNOMSSDIRFLMKENHYIKAEPTTLKFGTLAIEDDAELEIF 502
Oy 475 R--FTQT--GGVVISLQNCNVLISCYKNGACNSASNSITLKHIGLWLSLKGAIPLLVY 531
Db 503 NIPFTQNTSLLAGSGATLVGVKRG-----KLNTINLGVLLPILKKGKSPCIRV 554
Oy 532 EPTN-NSNNYATD-AAEFSL-----DVKLSLDDGNSPEYSTDTHALSSQPLSI 583
Db 555 NPODQNTGTGQIPSSSTISSTIPMIIFNGRSLYDENVESYOSMDLSRKAQOLLI 614
Oy 584 SEASDQRLSPD--DMDFSGLVNPHYGMQGLMTWMAKT-----ODPEPASSATITD 632
Db 615 ETTNGQLDSNMNSSLNLSLSPHYGYGLMTPTWITTTITLNNSSAPTSATISAE 674
Oy 633 POKANR-----FHRLLTLTWLPAGYVPSPKH 658
Db 675 QKTSSETFTPSNTTASIPNIKASGSGSANSNGEYTLTKHLLVYMAVGYIVDPIR 734
Oy 659 RSPLIANTL---WGNMLATESLAKNSAEUTPSDHPFGITGGGLGMVYODPRENHPGFH 715
Db 735 ROLLANSLSVHSGRNMHTGLNSL-----PDNSWFLQGAATTLFTKQKRSLSYHGS 787
Oy 716 MRSSGYSAGMIA--GOTFTSLKFSOTYTKLNERAKNNYSKNNY---SCGEMLFSLOE 770
Db 788 SASKGYTVSSQASGAHGHKFLSFGSSDKMKERETNRLSRYYLSALCFEHPFMD--- 844
Oy 771 GFLTKLVG--LYSGDNHCHHFFYTQGENLTSOGTFRSOTMGANFPL-----PKPFSGS 824
Db 845 ---RIALGAAACNCGTHNMRSPY--GTRKSSKGRKFSHTTLGASLRCELDRSMPIR--- 895
Oy 825 THILAPFLGALGIYSSHSFTEVGAPRFSFTKPLINLVPIGVKSGFMNATQORQAW 884
Db 896 -SMLTTPFOALFSTRTERPASIESCDLARLFLBEQAHRAVSPICIKAYSSDTPTLSW 954
Oy 885 TVELAYQVLYRQEPGIATOLLASKGIWFGSGSPSSRHAMSRYKISQOTQPLSMULTLFOY 944
Db 955 EMELAYQPTLYKKRPLNLTLLIQNNGSWTNTPLAKHSF-YGRGSHSLKFSHLKLPANY 1013
Oy 945 HGFYSSSTFCNLTGCEIALRF 965
Db 1014 QAEVATISTVSHYINAGALVF 1034

RESULT 9
G81722
polymorphic membrane protein E/F family TC0262 (imported) - Chlamydia muridarum (stra
C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: G81722
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heibelberg, J.F.; White, O.; Hicke
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: G81722
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1025 <TEF>
A:Cross-references: GB:AE002293; GB:AE002160; NID:97190298; PIDN:AAF39131.1; PID:9719
A:Experimental source: strain N199 (Mopn)
C:Genetics:
A:Gene: TC0262

Query Match 15.6%; Score 793; DB 2; Length 1025;
Best Local Similarity 27.0%; Pred. No. 7.5e-41;
Matches 269; Conservative 158; Mismatches 419; Indels 152; Gaps 35;
Oy 68 LOKTPREGAAVTTIDYLSFFDTQKEGIFYFAKNLTPESGAIYASPSPTVEIRDITIGPV 127

QY 421 FNKEADTGVSVYEGGATYNSADFHORNLOTKTPAPLITLSNGLCELDIEDAQLTVNRFTQTG 480
 Db 420 FNKEEDDQGVSVYEGGATYNSADFHORNLOTKTPAPLITLSNGLCELDIEDAQLTVNRFTQTG 479
 QY 481 GVSISGNCNAVLSCKKNGNGNSASNASITLKHIGLNLSSILKSGAEIPLLWVEPTNNSNNY 540
 Db 480 GVSISGNCNAVLSCKKNGNGSASNASITLKHIGLNLSSILKSGAEIPLLWVEPTNNSNNY 539
 QY 541 TADTAATFSLSDVKLSLIDDYDGNSPYESTDLTALSSQPMUISISEASCNOLRSDMDPFG 600
 Db 540 TADTAATFSLSDVKLSLIDDYDGNSPYESTDLTALSSQPMUISISEASCNOLRSDMDPFG 599
 QY 601 LNVPHYGQGLMTWGMATODPEPASSATITDPOKANFHRLLTLTWLPAGYVPSPKHRS 660
 Db 600 LNVPHYGQGLMTWGMATODPEPASSATITDPOKANFHRLLTLTWLPAGYVPSPKHRS 659
 QY 661 PLIANTLGNMLLATESLKNLSAEILPSDHPFWGJTGGLGMMVYODPRENHPGFHMRSSG 720
 Db 660 PLIANTLGNMLLATESLKNLSAEILPSDHPFWGJTGGLGMMVYODPRENHPGFHMRSSG 719
 QY 721 YSAGHIAQQTTFSLKFSQYTKLNERAKNNVSSKNSCGCEMLFSIQEGLTKLVGL 780
 Db 720 YSAGHIAQQTTFSLKFSQYTKLNERAKNNVSSKNSCGCEMLFSIQEGLTKLVGL 779
 QY 781 YSYGDNHCNHHYTGENTSGTFRSQTMGCAVFDDLPKMFSGHITLTAFLGALGYS 840
 Db 780 YSYGDNHCNHHYTGENTSGTFRSQTMGCAVFDDLPKMFSGHITLTAFLGALGYS 839
 QY 841 SLSHTEVGAYPRSESTKPLINLVPIGVKSGFENATQRPQAMVELAYQPVLYROEPG 900
 Db 840 SLSHTEVGAYPRSESTKPLINLVPIGVKSGFENATQRPQAMVELAYQPVLYROEPG 899
 QY 901 IATQLLAKGIMFGSGSSSRHMSYKTSQOTPLSMWLTILHFOYIGFSSSTFCNYLNGE 960
 Db 900 IATQLLAKGIMFGSGSSSRHMSYKTSQOTPLSMWLTILHFOYIGFSSSTFCNYLNGE 959
 QY 961 IALRF 965
 Db 960 IALRF 964
 RESULT 2
 F81722
 Polymorphic membrane protein E/F family TC0261 [imported] - Chlamydia muridarum (strain
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: F81722
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwin, M.; Neilson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255
 A:Accession: F81722
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-976 <TEF>
 A:Cross-references: GB:AE002293; GB:AE002160; NID:g7190298; PIDN:AAE39130.1; PID:g7190303
 A:Experimental source: strain N19g (Mopn)
 A:Genetics:
 A:Gene: TC0261

Query	March	Similarity	71.7%	Score	3648.5	DB	2	Length	976
Best Local	Similarity	70.5%	Pred.	No. 24e216					
Matches	689	Conservative	121	Mismatches	154	Indels	13	Gaps	3

QY	1	MKKAFFFLIGNSGLAREVPSRIETLMPNSVPDPTEKSLKISLTGDTNHNITNCYIDN	60
DB	1	MKKLFEPALIGSILGTREVPSSILKPLINPMTGTGFFPKVALLDTHNLDDYHLDN	60
QY	61	LYLIALIOLKTPNEGAAYITDYLSFFDTQKEGIFAKNLTPESGAIGYASPNSPTEVI	120
DB	61	LKCIILACIQRPIYEECAATVTVDYLGDSFTQKDCICFNLNLPESGVIGSPQNPITKI	120

OY	121	RDTIPVLENNNTCCRPFFSSNPNAVKN	IRREGCAHAONLVIINHHDVYGFPMKNFSYR	180
Db	121	NHTIDPVLENNNTCHRLMTJOTDPE	NEGKGAIRREGGALIHADVYISNNQNLVGFJKNAAYQ	180
OY	181	GGALSTANTPVVSENOSCFLFMDNIC	IQTNTPAGKGAIVAGTISNFSNENNCDDLEFFINMC	240
Db	181	GGALIS-ANTFAKKNKSSPLCLN	NSCJOTKTOGGKGAIVYSTRSCFENNKKDLPFIQNSG	239
OY	241	CAGGAIFSPDTCGLTGNKGNIVIV	NNBFCFKNVETASSEASDGAIXKTTPLDVTGNNGRIE	300
Db	240	CAGGAIFSPDTCGLTGNOGIVEY	SNHGFKNVNATNESDGGAIKATTRIDITJNNNSQIF	299
OY	301	FSDNITTKANGGAIYAPVYVL	LVNDGPTFFYFNINLANNGKGAIIYDGTSNSISADRAHIFN	360
Db	300	FSDNISRNFEGGAIHAPCLH	LVNGNPTFFYFNINLANHNGGAIYITIGTSTKISADHAHIFD	359
OY	361	ENIVTNVNTNANCTSTANBP	RNRNATVASSGCEILLAGSSONLJFYDPIEVSNACVSYS	420
Db	360	NNISANATNADSSSNTNPNHR	NATIMONSAGIEIGACKSONLJFYDPIQVYTNACVYD	419
OY	421	FNKEADOTGCVVSCATVVSAD	FHOJNLOTKPPARTLSLNGFCEIDHOLYTNREFTOG	480
Db	420	FNKDAISOTGCVVSCATVVSAD	ISQANLOTKPPARTLLSHGLCLIEDRAQOLYVNNNTOG	479
OY	481	GVSALGNGAVLSCYKNGACN	-----SASNATILKIHGILNLSLILKSGAEIPELLWE	532
Db	480	GIVALGNGAVLSSYOHST	DATQPTTTTDDASVTLNHIGLNPILKDGAEMPELLWE	539
OY	533	PTN-----USNNNTA	TAAFAFSLSDVKYLSLIDYGNPSPEYSTDTHALSSOPMUIS	584
Db	540	PISTTOGNTTYSOTPA	ASFSLNGATLSLIDEBGNSPYENTDLSRALYADPMAISEASD	599
OY	589	NOLRSDMDDFSGLANPIY	MOGLMTWGMAKTOPPEPASSATIIDPOKANRFRHTLLTWL	648
Db	600	NOLGESMDSF	SVNVPHYGMOGLMTWGMAKTEPPTTPATIIDPKANGFHRTLLTWL	659
OY	649	PAGYVPSRKHRSPL	IANTLNGNMLLATESIKNSAELTFSDDHPFWCJITGGLGIMNVYODPR	708
Db	660	PAGYVPSRKHRSPL	IANTLNGNMLLEFATEMLKSSGGELLDPRFWGITGGLGIMNVYODPR	719
OY	709	ENHNPFFHMRSSCVSAGM	IGOTHTFSLKPSQVYTTKINERYAKNNVSSKKNVSCOGEMIFSL	768
Db	720	KHPRPFHMHNTSGV	SACMIIITGNHTFESLRFSQSTTKLNERAKNVSSKKNVSCOGEMILSL	779
OY	769	DEGFLLTKLVLGY	SYGDHNCHEFYTOGENLTSOGITEFSOTMGAVEFDLLPKMFPGSTHIL	838
Db	780	DEGLMLTKLIGLY	SYGNHNSHMFYTOGEGEDLSSCEFHSHQFFCGAVEFDLLPKMFPGRTHIL	839
OY	829	TAPFLGALGIYSSL	SHFTVEGAYPRFSFKTPLINLVLPJGVYGSFMMANAQORQOANTVEL	888
Db	840	TAPFLGALGIYSSL	SHFTVEGAYPRFTTETPLINLVLPJGVYGSFMMANARQOANTVEL	899
OY	889	AOQPVLYROEP	IATOLASKSIGIWFSGSPSSHNAMYSIKISOOTPLSWLTLFHYNGFY	948
Db	900	AOQPVLYROEP	ISTOLJLACKGMWPFCHGSPASHNALAYKISOKTOLLRFATLDOLOHGY	959
OY	949	SSSTFCNTVNGEIALRF	965	
Db	960	SSSTFCNTVNGEIALRF	976	

RESULT 3
H72074
polymorphic membrane protein E/F family CP0286 [imported] - Chlamydomonas reinhardtii
C:Species: Chlamydomonas reinhardtii
C:Date: 23-Apr-1999
C:Accession: H72074
R:Kaltman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
N:Reference number: A72000
A:Accession: H72074


```

OY 608 -----WQGLTWGMAKTODPEPPASATITDPOKANRFRHRL----- 643
Db 545 GLEVRDNKIIYKLGANLRFENGAVTAGTVNPSAPAPPLTAAEPPLRASNSHLQSLSEG 604
OY 644 -----LLTWLPAGYVPSPKHRSPL----- 662
Db 605 LVYHNALALQDGMENVQHGTLRVSGGLQMRGDIITVPSGPIPIERPLAPLTOTEN 664
OY 663 -IANTLMGNMLLATESLK-----NSAELTPSDHPFWGITGGGLGMWYODPRENHPGFHR 717
Db 665 GIGLAGLGLDESDALQYKVGPMRLNPEKRYVTLILCPGIS-----FGQY-ANRTNYDVR 720
OY 718 SSGYSAGMIAGOT-----HTFSLKFSQYTKLNERAKANNVSSKNSCGEMLFSLQ 769
Db 721 VS-VERPWFVGQRGQTLFVLVGHGLHONSKIQLNLGQGLKTDPTVNOLEVPILGQGLEIAD 779
OY 770 EGFLLTKLVGLYSYGD-----HNCHEFTQGENLTSSQ-----PPRSOTMGCAVPE 815
Db 780 ESQVVKL-----GDGLQFDSQARITTAAPNVTETLTGTSSMANVYRGTATAGSKLF 833
OY 816 DLPMPFGSTHILTAFLGALGIYSSLSHTEVGAYPRSFSTKPLINVLVPIGVKGSFM 875
Db 834 -----LSLTRFSTGLVGNMTIDSNAS-----FGQY-INGHEQIECFILLDNOGNLK 880
OY 876 NATOPQAMTVE-----LAYOPVLYROE-----PGIATOLLASKGIMFSG----- 916
Db 881 EGSNLOGTWEVKNPNSASKAAFLPSTALYPLINESRSLPG-----KNLVGMQAILGGGCTCT 938
OY 917 ---SPSSRHMSYKISQO-----TOPLSWLTJLHFOY 944
Db 939 VIATLNGRRSNNPAGOSIIFVWQEPNTIARQPLNHNSTLTFSTY 981

RESULT 22
US-08-815-927-26
; Sequence 26, Application US/08815927
; Patent No. 6086890
; GENERAL INFORMATION:
; APPLICANT: MITTAL, SURESH K.
; APPLICANT: GRAHAM, FRANK L.
; APPLICANT: PREVIC, LUDVIG A.
; APPLICANT: BABIUK, LORNE A.
; TITLE OF INVENTION: RECOMBINANT PROTEIN PRODUCTION IN BOVINE ADENOVIRUS EXPRESSION
; FILE REFERENCE: 293102002101
; CURRENT APPLICATION NUMBER: US/08/815,927
; CURRENT FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: 08/164,294
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 26
; LENGTH: 983
; TYPE: PRT
; ORGANISM: Bovine adenovirus type 3
US-08-815-927-26

```

Query Match 2.9% Score 150; DB 3; Length 983;

Best Local Similarity 19.8% Pred No 0.0061; Matches 199; Conservative 108; Mismatches 392; Indels 304; Gaps 46;

```

OY 188 NTFVSENOGCLFMDNICIQTNTAGKGAIVAGTSNFSFNCDLFFINACCAGATP 247
Db 37 NCFV--ENGEATLAM--LVEKKPLTFDEKALTLGVRGIRINPAGILFETND--LASAVP 89
OY 248 SPICSLTGRNGNIVFYNNCKFVETASSASDGAIAKVTTRLDVYGNR-----GRI 239
Db 90 PLLAS--DEAGNYTL-----NMSDGLYTKDKMLAVKVGPGSLSDSNNALOYHTDGLT 140
OY 300 FFDNDITKNYGAIVAPVVT-----LVNDCGPTYFINN--IANNK 336
Db 141 VTDKVSIN-----TQAPLSTTSAGLSLLGLPSLHLEEFENLTVNTAGLOISNNALAVKY 196

```

```

OY 337 GGATIIDGTSNKSIS-ADRIALIFENIV-----TN--VTNANGTSTSNPPRRNAI 385
Db 197 GSGITVDAQNOLAASLGGGLSRDNKRTVYKAGPGLTITNOALITVATGGLQVNEGQQLQ 256
OY 386 TVASSSG-----EILLAG-----SSQNLIEFYDPIEVSNAGVSVFNKEADQGS--- 430
Db 257 NITAGGLNFANNSLAVELGSLHFPQGNNOVSLY-----GDGIDIRNRTVYRAGPQLR 312
OY 431 -----VVFSGATVNSADPFHORNLOTKTPAPLTLSNGF-----LCIEHQAULTVNR 476
Db 313 MLNHQLAVASG-----DGLVHSDTLRLKLSHGLTFENGAVRAKLGGLGTDGSRGV--- 366
OY 477 TOTGVVSLGNCAV--LSCYKNAGNSAS-----NASITLKHGLNLSLTKGAEL----- 526
Db 367 -RTGRGLRVANQVOYIFSGRGTAICTDSSLTLNIRAPLOFSGPALTAISLQSGPITYNSN 425
OY 527 -----PLTWEPNTN-----NSNNYADTAATFSLSDVKLSLIDYGNSPY 566
Db 426 NCTFGLSIGPGHMVQONRLQVNPAGLVFGQNNLVPLNADPLAISDSKISLSLQGLT-Q 484
OY 567 FSTDLTALS-----SOPULSIS-----EASDNQLRSDMDPFGLVN-----PHYG- 607
Db 485 ASNALTLISLNGLEFSNQVAIAKAGRLRFESSQALESSLVNGTLTDTVIRPLNGD 544
OY 608 -----WQGLTWGMAKTODPEPPASATITDPOKANRFRHRL----- 643
Db 545 GLEVRDNKIIYKLGANLRFENGAVTAGTVNPSAPAPPLTAAEPPLRASNSHLQSLSEG 604
OY 644 -----LLTWLPAGYVPSPKHRSPL----- 662
Db 605 LVYHNALALQDGMENVQHGTLRVSGGLQMRGDIITVPSGPIPIERPLAPLTOTEN 664
OY 663 -IANTLMGNMLLATESLK-----NSAELTPSDHPFWGITGGGLGMWYODPRENHPGFHR 717
Db 665 GIGLAGLGLDESDALQYKVGPMRLNPEKRYVTLILCPGIS-----FGQY-ANRTNYDVR 720
OY 718 SSGYSAGMIAGOT-----HTFSLKFSQYTKLNERAKANNVSSKNSCGEMLFSLQ 769
Db 721 VS-VERPWFVGQRGQTLFVLVGHGLHONSKIQLNLGQGLKTDPTVNOLEVPILGQGLEIAD 779
OY 770 EGFLLTKLVGLYSYGD-----HNCHEFTQGENLTSSQ-----PPRSOTMGCAVPE 815
Db 780 ESQVVKL-----GDGLQFDSQARITTAAPNVTETLTGTSSMANVYRGTATAGSKLF 833
OY 816 DLPMPFGSTHILTAFLGALGIYSSLSHTEVGAYPRSFSTKPLINVLVPIGVKGSFM 875
Db 834 -----LSLTRFSTGLVGNMTIDSNAS-----FGQY-INGHEQIECFILLDNOGNLK 880
OY 876 NATOPQAMTVE-----LAYOPVLYROE-----PGIATOLLASKGIMFSG----- 916
Db 881 EGSNLOGTWEVKNPNSASKAAFLPSTALYPLINESRSLPG-----KNLVGMQAILGGGCTCT 938
OY 917 ---SPSSRHMSYKISQO-----TOPLSWLTJLHFOY 944
Db 939 VIATLNGRRSNNPAGOSIIFVWQEPNTIARQPLNHNSTLTFSTY 981

```

RESULT 23

US-08-614-377A-7
; Sequence 7, Application US/08614377A
; Patent No. 5976864

```

; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF
; TITLE OF INVENTION: HETEROLOGOUS
; TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson PC
; STREET: 225 Franklin Street

```


Mon May 7 09:06:32 2001

us-09-677-752-2.rai

Page 12

US-08-617-697-2

Query Match	3.5%	Score 176.5;	DB 2;	Length 1536;
Best Local Similarity	20.68;	Pred. No. 6.9e-06;		
Matches 134;	Conservative 98;	Mismatches 236;	Indels 181;	Gaps 28

[illegible]

```

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1
```

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/08/808,599A
3 FILING DATE: 28-FEB-1997
4 CLASSIFICATION: 424
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: US 08/317,522
7 FILING DATE: 04-OCT-1994
8 PRIOR APPLICATION DATA:
9 APPLICATION NUMBER: US 08/439,818
10 FILING DATE: 12-MAY-1995
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Campbell, Cathryn A.
13 REGISTRATION NUMBER: 31,815
14 REFERENCE/DOCKET NUMBER: P-LA 2255
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (619) 535-9001
17 TELEFAX: (619) 535-9949
18 INFORMATION FOR SEQ ID NO: 24:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 1160 amino acids
21 TYPE: amino acid
22 TOPOLOGY: linear
23 MOLECULE TYPE: protein
24 US-08-808-599A-24

```

Query Match	3.4%	Score 174;	DB 4;	Length 1160;
Best Local Similarity	23.7%	Pred. No. 7.1e-06;		
Matches 132;	Conservative 62;	Mismatches 200;	Indels 164;	Gaps 25

OY	68	LOKIPNECAATVITIDVUSLFFPDTOKEGIYFAANILPESGALIGVYSPPMVEIHDITGPV	127
Db	544	LNSASFGSALSTASFGCVLNGRAGFGGALINTNATGCVLNGSAGCGGAGAMNTATFGGA	603
OY	128	IFENNTCCRP-----TSSNPAAVNKIRE-GGAIHAONLEYINHHDVYEFMKNFYSVRG	181
Db	604	LNSN-----AGFGAISTSTFNFGALLNSAGCFGAMNTSASFGVLNNSAG-----G	651
OY	182	CAISTANTFVYSEMOCSFLFMDNICIOTNTGKKGALYAGTSNSESFNCDLFFINNACC	241
Db	652	GAINTSANFGGA-----LTSNAGFGGAI--STSASFGG-----ALNNSAG	689
OY	242	AGCAIFSPICISITGCRGNI VFYNNRCFKNVEPASSPASDGAIVVTRILDOVGNRGRIFF	301
Db	690	FGGAL-STSASFGGALLNSAGFGG-----AISTMSFGGAI-----	724
OY	302	SDNITKNGGAIYAPVYVLVNGPETYFINNIANKNGAIYIDGTSNKIS-----	351
Db	725	---SNSPDDGCGAF---STSVGFCGLLNTTDCGSHNSNISFGSAPPTSVSFGSGHSTNLC	778
OY	352	---ADRHAIIF---NENIV---TIVTNAANTSTISANPRRNAT---VASSGGEILLG	397
Db	779	FGGAPSTSLCGSASNTNLCFGGSSNTSCFGATSANNEGHSISFGNGLSTSG---FG	835
OY	398	AGSSGNLIFYDPIEVS-----NAGSVSFNKE-ADQTC-----SVYFSGATYNSADFH	444
Db	836	NGLGTSAGFDSLSLGTSTGFGSGLSPSASFNGLGTSTGFGGGLGTSTDFSGLNHNADF	895
OY	445	-----QNNLOTKTP-----APLTSNGFLICIEDHAQTV	472
Db	896	CGLGNASACFNGCLNNTNTPFGELCTISAGFCGJGLSSSTFSFGAGLVTSDF-----AG	946
OY	474	NRTOTGGVNSLGNCAVLSCYKNGAGNSASNASITLKHIGLNTLSILKSGAEIPLWVER	533
Db	947	NLGNTGTGGGTLGTGAGFSVSLNN-GNGFCGMPNASFNRGLNTIIGFSG-----	995
OY	534	TNNSNNTAD--TAATFS	549
Db	996	SNTSGFTGEPRTGSSFS	1013

RESULT 17
US-08-728-470-10

DB 821 KGIYAKKNITFE-GGNITFGSKRAVT-EIE---GNVTINNANVTLLIGSDFDNHOKPLTI 875
140 -----SSNPNAAVVKIKREGGAIHAONLYINHHVDVCGPMKFSYVRCGALSTANTFVY 192
DB 876 KKVJINSNLTAGNITVNIAG-----NLTVESNANFKA-ITNFTNVGGLFDNKGNSNI 929
QY 193 SENOSCFLEMD-----NICIOTNAGKGAIVAGTSNFESENNCDLFEINNACCA-----G 243
DB 930 SIAGGARFKDIDNSKNLSTITSSSTYRTIISG---NITNKGDLNITEGSDTEMOIG 986
QY 244 GAIFSPICSLTGKRGIVYNNRCFK---NVEFASSEASDGC-----AIKVTTLDDVT 293
DB 987 GDVSOKEGMLTISDDKINTKQITIKAGVDGENSDSDATNANMLIKTKELKLTODLNI 1046
QY 294 G-NRGRIEFSNDITKNYGAIVAPVTLVDNGPTYFINNIANNKGAIIYDGTNSKISA 352
DB 1047 GFNKAETTKDGSDLTIGNT-----NSADGTNNAKVTENQYKDSKISA 1089
QY 353 DRHAIIFENENIVNTNANGTSTANP-----PRRAITVASSSGEILL 396
DB 1090 DGHKVTLSKSVETSGSNNTEDSSDNAGLTIDAKNVTNNITSHKAIVSISATSGEITT 1149
QY 397 GAGSSONLI-----FYDPIEVNAGVSVFNKKA-----DOTGSVVF-----SGA 436
DB 1150 KTGTTINATGNVEITAOGTGSLIGIESSGSVTLTATEGALVANSISGNTVYVTANSKA 1209
QY 437 TVNSADPHORNLQTKTPAPL-----TLNCGFLICEDHAOLTVN-----474
DB 1210 LTTLASTIKGTESVTTSSGSDIGGTISGTYEVKATELTLTOSNKAIKATTEANVT 1269
QY 475 -----RFTGTGGVVSILGNAGVLSCKNAGNSASNASITLK---HI-----512
DB 1270 ATGTIGTISGNTVNTNANGDLTVNGAEINATEGAATLTSSGKLTEASSHITSAGK 1329
QY 513 GLWLSILKSGAEIPLWVEPTNNSNNYTDATFSLSDVKLSLD-----559
DB 1330 QVNLISA--QDGS-----VAGSINAANVTLTNTGT--LTVKGSININATSGTLIVINAKDA 1379
QY 560 -----DYGNSPYESTDLTHALSSQPMLSISEASDNOLRSDMDPSGLNV 603
DB 1380 ELNGALGN--HTVNVATNANGSGSVIATITSSRVN-ITGDILITINGINI 1425

RESULT 13
US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berekteser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS-VG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match 3.5% Score 176.5 DB 2: Length 1536:
Best Local Similarity 20.6% Pred No 6.9e-06:
Matches 134: Conservative 98: Mismatches 236: Indels 181: Gaps 28:

QY 92 EGIYFAKNLTPFSGAIGVAPSPVTEIRDTIGYIFENNNTCC-----RPFT- 139
DB 821 KGIYAKKNITFE-GGNITFGSKRAVT-EIE---GNVTINNANVTLLIGSDFDNHOKPLTI 875
QY 140 -----SSNPNAAVVKIKREGGAIHAONLYINHHVDVCGPMKFSYVRCGALSTANTFVY 192
DB 876 KKVJINSNLTAGNITVNIAG-----NLTVESNANFKA-ITNFTNVGGLFDNKGNSNI 929
QY 193 SENOSCFLEMD-----NICIOTNAGKGAIVAGTSNFESENNCDLFEINNACCA-----G 243
DB 930 SIAGGARFKDIDNSKNLSTITSSSTYRTIISG---NITNKGDLNITEGSDTEMOIG 986
QY 244 GAIFSPICSLTGKRGIVYNNRCFK---NVEFASSEASDGC-----AIKVTTLDDVT 293
DB 987 GDVSOKEGMLTISDDKINTKQITIKAGVDGENSDSDATNANMLIKTKELKLTODLNI 1046
QY 294 G-NRGRIEFSNDITKNYGAIVAPVTLVDNGPTYFINNIANNKGAIIYDGTNSKISA 352
DB 1047 GFNKAETTKDGSDLTIGNT-----NSADGTNNAKVTENQYKDSKISA 1089
QY 353 DRHAIIFENENIVNTNANGTSTANP-----PRRAITVASSSGEILL 396
DB 1090 DGHKVTLSKSVETSGSNNTEDSSDNAGLTIDAKNVTNNITSHKAIVSISATSGEITT 1149
QY 397 GAGSSONLI-----FYDPIEVNAGVSVFNKKA-----DOTGSVVF-----SGA 436
DB 1150 KTGTTINATGNVEITAOGTGSLIGIESSGSVTLTATEGALVANSISGNTVYVTANSKA 1209
QY 437 TVNSADPHORNLQTKTPAPL-----TLNCGFLICEDHAOLTVN-----474
DB 1210 LTTLASTIKGTESVTTSSGSDIGGTISGTYEVKATELTLTOSNKAIKATTEANVT 1269
QY 475 -----RFTGTGGVVSILGNAGVLSCKNAGNSASNASITLK---HI-----512
DB 1270 ATGTIGTISGNTVNTNANGDLTVNGAEINATEGAATLTSSGKLTEASSHITSAGK 1329
QY 513 GLWLSILKSGAEIPLWVEPTNNSNNYTDATFSLSDVKLSLD-----559
DB 1330 QVNLISA--QDGS-----VAGSINAANVTLTNTGT--LTVKGSININATSGTLIVINAKDA 1379
QY 560 -----DYGNSPYESTDLTHALSSQPMLSISEASDNOLRSDMDPSGLNV 603
DB 1380 ELNGALGN--HTVNVATNANGSGSVIATITSSRVN-ITGDILITINGINI 1425

RESULT 14

```

Db 648 TRKNTSYMTSDSHSNWNSALNLETGANFTFKIYSSNSKGLTTOYRSSAGV----- 699
Oy 135 CRPTSSNPNAAVNRIREGGAIHAQ-----NLXI 163
Db 700 --NFNGVGNMMSFN-LKEBAKVNFKXPNENMNTSKPLIRFLANTTATGCGSVFFDIYA 756
Oy 164 NHN-----HDVGFMKNFYSYRGAL-----STANTFVYSEMSOCL 200
Db 757 NHSGGAELKMEINISNANFTLN-SHYRGDAFKINKDLINATNSNFSLRQTKDFY 815
Oy 201 --FPMNICIOT-NTAGKGAIIAGTSNSFESNNCLFF-----INNA----- 239
Db 816 DQYARNAISTNTNISTLGGNVTLLGGONSSSTTGNTIEKANVTLEANNAPNOQIRDR 875
Oy 240 -CCAGGAFSPICSLTGN-----RGNIVFNNRCFK-----NVEASEASD 280
Db 876 VIKLSLVLNGLSLJLGEADIKGNLTISESATPKCKTDTLNTGNTFNNGTAELNITQ 935
Oy 281 G-----GAIVTTRL-----DVTGNRGRIFFSD----- 303
Db 936 GYVKLGNTNDDDLNITTHAKNRORSIIIGDIIINKKSLINTDSNDAEIQIGNISQKE 995
Oy 304 -----NITK-----NYGAIYAPVTVLD----- 322
Db 996 GMLTSSDKINITKQITTKGIDGEDSSDATSMNLTKTKELKLELISGFNKAEI 1055
Oy 323 --NGPTYFINNIANNKGA-----IYIDGTSNKSISADRHAIIFNENIYNTVNGNST 375
Db 1056 TAKDQDRLTIGNSDNGSCAEKATYTFNNVKDSKISADCHNTLNSKVYTTSSNGRESN 1115
Oy 376 SANPRRRAITVA-----SSGCEILLGASSONLIFYDPIEVSNAVY 417
Db 1116 SOND--TGLITTAKNVEYNKDIITSLKTYNITASEKVITTAGST-----INATNGKA 1164
Oy 418 SVSPFKNEADOTGSYVFCATVN-SADPHORNLOTKTPAPLTIJNSGFLCIEDHAQL----- 471
Db 1165 SIT-RTKGDISCT--SCNTVSATV--DLTKSGKIEKSGCANVTSATGTIGCTI 1218
Oy 472 --TVNRFOTGTVSLGNGAVLSCKYKNGAGNSASNAITLKHIGLISLKSAGIEPL 528
Db 1219 SCNTVN-ITANAGDLTVGCAGIENATEGAATLTAICNTLTTE-AG--SSITSTGOVVL 1273
Oy 529 L-----WVEPTNNSNNTYADTAATFSL--SDVKL--SLIDU-----YGNSPYSTDIT 572
Db 1274 LAQNGSINGSIANAANTLTNTGTLTVAAGSDIKATSGTLVINAKDAKLMDASGDSDEV- 1332
Oy 573 HAISSOPMLTISEASDN--QLKSDMDFSGLNV 603
Db 1333 NAVNASGSGSVTAATSSSVNITGDLNTVNGLNI 1365

RESULT 10
US-08-038-682-2
Sequence 2, Application US/08038682
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GENE II, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Matlare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,682
FILING DATE: 16-MAR-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single-
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-682-2

Query Match 3.5% Score 176.5 DB 1: Length 1536:
Best Local Similarity 20.6% Pred. No. 6.9e-06:
Matches 134; Conservative 98; Mismatches 236; Indels 181; Gaps 28:

Oy 92 EGIYFAKNLTPEPSGAGIVASPSPTVEIRDTIGPIFENNTCC-----RPT- 139
Db 821 KGIYAKKNITTFE-GGNITTFGSKAAT-EIE--GAVTINNANVTLLIGDFPNHOKPLTI 875
Oy 140 -----SSNPAAVNKIREGGAIHAQNLVINHNDVYGFKNFYSYRGCAISTANTFVY 192
Db 876 KRDVLIINSGLNLAGGNIYVIA-----NLTVESNANFKA-ITNFENNQGLDNGNSMI 929
Oy 193 SNNOSCLFMD-----NICIOTNTAGKGAIIAGTSNSFESNNCLFFINNACCA-----G 243
Db 930 SJAKGARPKRIDNSKNKSIITTNSSYVHTIISG--NITNNGDLNITNEQSDTEMOIG 986
Oy 244 GAIFPICSLTGNRCNINIVYNNRCFK--NVEASEASDGC-----AIKVTTRLVYT 293
Db 987 GIVSKEGKGLTISDKINITKQITTKAGVDGNSDSTATNANLTKIKELKLTODLNI 1046
Oy 294 G-NKGRIFESDNIITKNYGAIIYAPVTVLDNGPTYFINNIANNKGAIIYIDGTSNKSITA 352
Db 1047 GFNKAEITAKDSDLTIGNT-----NSADGTNKKVTFNQVKDSKISA 1089
Oy 353 DPHAIIFNENIYNTVNGNSTSANP-----PRNAITVASSGCEILL 396
Db 1090 DGHKVTLSHKVETSGSNNTEDSSDNAGLTIIDAKNVYNNNITSHKAVSISATGEIIT 1149
Oy 397 GAGSSONLI-----FYDPIEVSNAVGSVPNKA-----DQGSYVF-----SGA 436
Db 1150 KTGTTINATTVETTAOTGSLIGLIESSGCVTLTATGALAVSNIGNTVTVANGSA 1209
Oy 437 TVNSADPHORNLOTKTPAPL-----TLNGLFLCIEDHAQLTVN----- 474
Db 1210 LTLTGSTIKGESVTTSSQSDIGITISGCVETKAKTESLTQSNKIKATGTGANT 1269
Oy 475 -----RFTQGTGVYSLGNGAVLSCKYKNGAGNSASNAITLK--HI----- 512
Db 1270 ATGTIGGTTISGNTVNVTVNAGDLTVNGAENATEGAATLTITSSGKLTTEASSHITSAG 1329
Oy 513 GLNLSLKSGAIEPLWVEPTNNSNNTYADTAATFSLSDVKLSID----- 559
Db 1330 QVNLISA--QDGS-----VAGSINANVTLTNTGT--LTVKGSININATSGTLVINAKDA 1379
Oy 560 -----DYGNSPYESTDLTHALSSOPMLTISEASDQLKSDMDFSGLNV 603
Db 1380 ELNGAALGN--HTVNAVNTANSGSVIATTSRVN-ITGDLITINGLNI 1425

RESULT 11

```

Db 1219 SCNTVN-VTANAGDLTVGNGCAEINATEGAATLTATGNTLITE-AG---SSITSTKGQVDL 1273
OY 529 L-----WVEPTNSNNYTDATATFSL---SDVKL---SLIDD-----YGNSPYESTDLT 572
Db 1274 LAONGSIAGSINAANVTLLTGTTLTVAAGSDIKATSGTLVINAKDAKLGASDGSSTEV- 1332
OY 573 HALSSQPMLSISEASDN--QLRSDMDPFGCLNV 603
Db 1333 NAVNASGGSVTAATSSSVNITGDLNTVNGCLNI 1365
RESULT 7
US-08-469-880-4
Sequence 4, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:V9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO. 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-880-4
Query Match 3.5%; Score 177; DB 2; Length 1477;
Best Local Similarity 19.4%; Pred. No. 5.8e-06;
Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

OY 135 CRPTSSNPAANVKKIREGCAIHAQ-----NLXI 163
Db 700 --NENGVMGNSFN-LKEGAKVNEFKLPENMNMTSKPLPIFLANITATGGSYFEPIYA 756
OY 164 NHN-----HDVVGFMKNFSYVRCGAI-----STANTFVSENGSCFL 200
Db 757 NMSGRAELKASEINISGANFTLN-SHVRGDDAFKINKDLITATNSNFSLRQTKODFY 815
OY 201 --FMDNICIOT-NTAGKGAIAVAGTSNFSFESNCDLEF-----INNA----- 239
Db 816 DGVAARNAINSTYNTISILGANTVLLGGQNSSSITGNITEKANANTLEARNAPNOQNTDR 875
OY 240 -CCAGGAIFFPICSLTGN---RCNIVFYNNRCK-----NWTASSEASD 280
Db 876 VIKLGLVNGSLSLTGENADIKGNLTISESATFKGTRDTLNTGNETNGTAEINITQ 935
OY 281 G-----GAIKVYTRL-----DVTGNGRGRIFPSD----- 303
Db 936 GVVKLGNTNDGDLNITTHAKRNORSIIIGDIIKKGSLNITDSNNDAEIOIGCNISQKE 995
OY 304 -----NITK-----NYGAIYAPVYLYD----- 322
Db 996 GNLTISSUKINITKOITIKKGIDGEDSSDATSANLTIKTKELKLFEDLSIGFNKAEI 1055
OY 323 --NGPTYFINNIANNKGA-----IYIDGTSNKSISADRAHIIENENIVTVNANGTST 375
Db 1056 TAKDGRDLTIGNSNDGNGSGAEKIVTFNNVKDSKISADGHVYTLNLSKYKTISSNGREGSN 1115
OY 376 SANPRRNAIYVA-----SSSGELLGAGSSONLIFYPIEVSNGV 417
Db 1116 SPND---TGLITTAKNVEYKNDITSLKTVNITASKEVTTAGST-----INANTNKA 1164
OY 418 SVSEFKKADOTGSYVFSGATVN-SADFHQRNLOTTPAPLTLNSGFLCIEDHAOL----- 471
Db 1165 SJT-TKTGDISGTI--SCNTYVSATV---DLTTSKSGKIEAKSSEANVTSGTIGCTI 1218
OY 472 ---TVNFTOTGCVVSLGNGAVLSCYKKGAGNSASNASITLKHIGLNLSSILKSAGAEPL 528
Db 1219 SCNTVN-VTANAGDLTVGNGCAEINATEGAATLTATGNTLITE-AG---SSITSTKGQVDL 1273
OY 529 L-----WVEPTNSNNYTDATATFSL---SDVKL---SLIDD-----YGNSPYESTDLT 572
Db 1274 LAONGSIAGSINAANVTLLTGTTLTVAAGSDIKATSGTLVINAKDAKLGASDGSSTEV- 1332
OY 573 HALSSQPMLSISEASDN--QLRSDMDPFGCLNV 603
Db 1333 NAVNASGGSVTAATSSSVNITGDLNTVNGCLNI 1365
RESULT 8
US-08-728-470-4
Sequence 4, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Maltare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/038,682
 FILING DATE: 16-MAR-1993
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: BERSKRESSER, JERRY W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-293
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1477 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-038-682-4

Query Match 3.5%; Score 177; DB 1; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
 Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

OY 37 KESLSNKSISLTCG-----THNLNLCYLDNLRILAILOKTPNEGAAYTID- 82
 DB 603 KDFRANNNVSLNCTGKGLNISSVNNLTNHLSC-----TINISGNITINOT 647
 OY 83 ---YLSFDTOKEGIYFAKNLTPESG---AIGYASPNSP--TVEIRDTIGPIEFENNTC 134
 DB 648 TRKNNTSYMOTSHDSHMVNSALNLETGANFTFKIYISNSKGLTTOYRSSAGV----- 699
 OY 135 CRRPTSSNNAVNKIREGCAHAQ-----NLXI 163
 DB 700 --NENGVMGMSFN--LKEGAKVNEFLKPNENMNTSKPLPIRFLANTATGSGSVFEDITA 756
 OY 164 NNN-----HDVVGFMKNFSYVVGAI-----STANTPVVSENOSEFL 200
 DB 757 NNSGGAELKMEINISNCANFTLN--SHVRCDDAFKINKDLTINATNSNFSLRQFKDFEY 815
 OY 201 --FMDNICIOT--NTAGKGAITYAGTSNFSFNKDLFF-----INNA----- 239
 DB 816 DGYARNAINSTNISILGONVTLGGONSSSITGNITIEKAAVNTLEANNAPNOQIRDR 875
 OY 240 --CCAGAIPIPSICSLTGN---RCNIVFYNNRCFK-----NVTASSEASD 280
 DB 876 VKLKSILVNGSLSLTGEVADIKGNLTISESATPFKKTDTLNTIGNFNGIAELINITQ 935
 OY 281 G-----GAIVYTRFL-----DYTGNGRGRIFFSO----- 303
 DB 936 GVVKLGNVTNDDLNTTTAKRNORSIIIGGDI INKKGSLNITDSNDAEIQICGINSOKE 995
 OY 304 -----NITK-----NYCGAIYAPVTVYD----- 322
 DB 996 GNLITSSDKINITIKOTIKKIGIDGEDSSDASNNANLITIKELKLTEDLSISGFNKAET 1055
 OY 323 ---NGPTFYNNANNGKA---IYIDGTSNKSISADRAHAIIFENINITYNTNANGTST 375
 DB 1056 TAKDGDLLTIGNSNDGNSGAEKATVFNNAVKDSKISADGHVNTLNSKVTSSNGGRESN 1115
 OY 376 SANPPRRNAITVA-----SSSGEILLGAGSSONLIFYDPIEVSNAGV 417
 DB 1116 SDMD---TGTLITAKNVEVKNKDLTSLKVTINITASEKVTYTTAGST-----INATNGKA 1164
 OY 418 SVSFNKEADQOTGVSFSGATVN--SADFIQRNLOTKTPAPLTLNSGFLCIEDHQL----- 471

DB 1165 SIT--TKTGDISGTT--SGNTVSVSATV---DLTKSGSKIEKSGEAVNTATGTIGCTI 1218
 OY 472 ---TVNRFOTGTGVSILNGAVLSCYKNGACNSASMAITLKHIGLNLSTILKSGAETPL 528
 DB 1219 SGNTVY--VTANAGDLYVNGAELINTEGATLTATGNLTTE--AG---SSITSTKGOVDL 1273
 OY 529 L---WVEPTNNSNNTYADTATFSL---SDVKL---SLID-----YGNSPYESTDLT 572
 DB 1274 LAQNSIAGSINAAVNTLTGTLTTVAGSDIKATSGTLVINAKDAKLNCDASGDSTEV- 1332
 OY 573 HALSQPMLSTISESDN--QLRSDMDSEGLNV 603
 DB 1333 NAYNAGSGSVTAATSSVNTIGDLNTVNGLNI 1365

RESULT 5

US-08-302-832-4
 Sequence 4, Application US/08302832
 Patent No. 5603938

GENERAL INFORMATION:
 APPLICANT: Barenkamp, Stephen J
 TITLE OF INVENTION: High Molecular Weight Surface Proteins
 TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Shoemaker and Matlare, Ltd.
 STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
 CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.

ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/302,832
 FILING DATE: 16-SEP-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9205704.1
 FILING DATE: 16-MAR-1992

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US pct/us93/02166
 FILING DATE: 16-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Berskresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-404
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1477 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-302-832-4

Query Match 3.5%; Score 177; DB 1; Length 1477;

Best Local Similarity 19.4%; Pred. No. 5.8e-06;
 Matches 158; Conservative 115; Mismatches 244; Indels 296; Gaps 40;

OY 37 KESLSNKSISLTCG-----THNLNLCYLDNLRILAILOKTPNEGAAYTID- 82
 DB 603 KDFRANNNVSLNCTGKGLNISSVNNLTNHLSC-----TINISGNITINOT 647
 OY 83 ---YLSFDTOKEGIYFAKNLTPESG---AIGYASPNSP--TVEIRDTIGPIEFENNTC 134

Matches 190: Conservative 135; Mismatches 434; Indels 244; Gaps 36;

OY 41 SNKISLTGDT-----HNLNTCTIDNLRILALILKOTPNNGCAVTTIDVLSFPTQEG 93
Db 602 SKTILNADVAINELVYENNGSVOLNHNHTYLTITKINAACQOILVADPLNTTTLADG 661
OY 94 IYFAKNLTPESGGAIGVAPSPSYEIRDTIGPVIFENNTCCRPFTSSNPNAVKIREG 153
Db 662 TNLGSAENPLSTIHFATYAANDSI-LVVGKGVNLXANN-----ITTDANVAGSLHFRSG 715
OY 154 GAI-----HAQNLVINHNHDVYEFKMFESVRCGALSTANTFVSNQSC----- 198
Db 716 GTSIVSGVGGQGHKLNNLLIDNGTIVKFLGDTTFNGGTRKIEGKSILOISNNYTDHVE 775
OY 199 -----FLPMNICTIOTNTAGCGAIYACTSNSFESNCDLFF--INNACCGCAIF 247
Db 776 SADNTGTLFVNTDPI---TTLNKGAYFGVLKQVYIISGPNIFENIGVNGIVHGIAA 832
OY 248 SPICSLTGRCIVIVYNNRCFKNVETASEASDG-----GAIKVTTRLDVGNRGRI-- 299
Db 833 NSISFEMASLGSLFLPSTPLDVLITSTVGNQIVDNFNAPIVVYSGIDSMINNGIIG 892
OY 300 -----FFSDN-ITKNVCGAIYAPVTVLVNCPYFPIINNIAKNGAIYIDGTSNS- 348
Db 893 DKKNIIALSLGSDNITVN-ANTLYSGIRTKNNOGTITLGGMPNPGITVYGLGLENGS 951
OY 349 -KI-----SADRAHIIENENIVTVNTNANGSTSNAPRRRAIVASSSGE- 393
Db 952 PLKQVTFETTDYNNLGSIIANNVYTIINDVTLTGIGIAGTDFDAK-----ITLGSVNGNA 1005
OY 394 -----ILLGAGSSONLIEFYDPIEVSNAQ-----VSSEFNKEADQTSVYV 432
Db 1006 NVRFVDSTSDPSRMIVATQANKVTYVIGNALVSHIGSLDTPVASVRT--GNDGAGL 1063
OY 433 FSGATVNSADFHORNLOTKTPAPLTLNGLFCIEDHAQLTVNRFOTGCVVSLGCAVLS 492
Db 1064 QGVIYQONIDFGTYNL-TILNSVNLGGGTTAINGEIDLTLNLIIFANGSTMGDTNIS 1122
OY 493 CYKNGAG-----NSASNASITLK-----HIGLNLSSILKSGAKIPLLM 530
Db 1123 TJLVNSSGNIGVYIAEDAOVNATTTGTTTIIKIQDNANANFSGTQAVYIIOGGRF- 1178
OY 531 VERTNNSNNTATATAFSLSDVKLSLID-----YGNSP 565
Db 1179 -----NGTLGAPFAVATGSNIFVKYELIRDSNODVYLTRTNDVLANVYTTAVGSAIANAP 1233
OY 566 YESTDLTHALSQPLSLISEASDNLSDMDFSGLNVPHYGMOGLMTWGMATQDDEPA 625
Db 1234 GVSQONISRCLES-----TNTAAYNNMLLAKDPDVAFTV-----GAITDITSAV 1278
OY 626 SSATITDPOKANRHFRTLLTWLPAGVYSPKRRSPLIANTIMCNMLLATESLKNSELT 685
Db 1279 TJVLNLTQKTOD-----LLSNRLCTLRVLSNAETSDVAGS---ATGAVSSGOEAE 1326
OY 686 PS-----DHPMGITGGGLMMVYODPRENIHGFHMRSSGSAGM--IAGOTHTFSLAESQ 739
Db 1327 VSTGVMAKPTTINI-----AEDKKGAGIAGYAKATTGGVYVGDITLASDMLMGAIGI 1378
OY 740 TYTKL-NERYAKNNVSS-----KNVSCOGEMFLSLQEOFLITKLVLGLXSY 783
Db 1379 TKTDIKHODYKKGDKDTINGLSFLYGSQOLVKNEFAQNSIFTLNK-----V 1426
OY 784 GDHNCHEFTYOGENLTQSO---GTFRSOTMGCAVFF--DLPRKPGSGHTLILAPLGAIGI 838
Db 1427 KSKSORFFESNGKMSKQIAAGNYDNMTPGNLIIFGYDYNAMP-----NVJVTPMAGISYL 1482
OY 839 YSSLIHTEVG---AYPSFSTKTLPLINVLVPIGVKGSFMMAT 878
Db 1483 KSSNENKETGTJYANKRIJNSKFSNRVDLIYCAVAVAGSTVNT 1525

Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302.832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Baretteser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

Query Match 3.5%; Score 177.5; DB 2; Length 1338;
Best local Similarity 21.4%; Pred. No. 4.5e-06;
Matches 154; Conservative 93; Mismatches 286; Indels 187; Gaps 33;

OY 75 GAAVTTIDYLSFPTQEGIYFAKNLTPESGGAIGVAPSPSYEIRDTIGPVIFENNTC 134
Db 546 GGNITIRQVEGTDSRVNKGVAANKNIT-FKGNITTFESOKA-TIEIK--GNVTINKNTN 600
OY 135 CRPFTSSNPNAVK-----IREGAI--HAQNLVI-----NNHD 168
Db 601 A---TLGANTAEKMSPLNINAGVINGNLTITACISIIINAGNLVSGANLOAITNTFN 657
OY 169 VVGFM-----KNESYVRCGA-----ISTANTVVEENOSCFLEPMNICTIOTITAKGAIY 219
Db 658 VAGSFIDNAGSNISIRAGAKFKDINNTSSLNTITNSDTTY--RTIIKGNISKSG-- 711
OY 220 AGTSNFSFNNCDLFFINNACCA-----GGAIFSPICSLTGRCNIVFVNNCFK-NVE-- 272
Db 712 -----DLNIIIDKSDAEIQIGNISQKEGNLTISDKVNTINOTITAKYEGG 759
OY 273 -TASEASDGAIVYTRLDVATGNRGRIFFSDNITKNYGAIVAPVTVLVNCPYFPIINN 331
Db 760 RSDSEANLTIOTKELKLAGD-----LNISGFNKAIEIT--AKNGSDLTIGN 806

```

OY 590 QLRSDMDPSGLN-----PHYGQGLMTWMAK-TODPEPASATITDPQK 635
Db 585 -----FKGATVTKTGFPCDEIATPSHYGQKMSYTWMSRPLIPADCGPCGSPSPS 636
OY 636 ANRFRHTLLTW-----LPAGVPSPKHRSPLIANTLMGNMILLATESLKN5A-----EL 684
Db 637 AN---TLVAVWNSDTLVRSYIILDPERYGEIVNSLW-----ISFLGNQAFSDILODV 686
OY 685 TPSDHPFMGITGGIGMMVYODPRENHPGFHMRSSSGYSAGMIACOTN--TFSLKFSQTYT 742
Db 687 LLIDHPGLSTAKALGAYEHNTPRGCHGFGSGRYGQYQALSMNYTDHTTLGLSFGQLYG 746
OY 743 KLNERYAKNNVSSKNY--SCQGMFLSLQEGFLTKLVGLYSYGDHNCHEFTYQGENL-T 799
Db 747 KTNANPYDSRCSQWYLLSFCQFPVIVOKSEALISMKAAYGYSKNHLNTYLRPDKAPK 806
OY 800 SOGTFRSQTMGCAVFFDLPMKPFQSTHILTA-----FLGALGIYSSLSHFTYGA 850
Db 807 SOGQWHNNSYVLLISAE--HPFLNMCLITRPLAOAMDLSGFTISAEFLGQWOSKFTETGD 863
OY 851 YPRSFSTKTPLINLVPIGVKGSFNNATOR--POANTVELAYQPVLYRQEPGIATQULASK 909
Db 864 LQRSFS-RGKGYNVSLPIGCSQWFTPFKAPSTLTIKLAYKPDYRVNPHNIIVTVVSNQ 922
OY 910 GIWFGSGSPSSRHMSYKISOOTOPLSMLTLHFQYHGFTSSSTFCNY 956
Db 923 ESTSISGANLRRHGLFVQI-HDVVDLTEDTQAFLLNYTFDGKNGFTNH 968

```

Search completed: May 6, 2001, 19:23:26
 Job time: 401 sec

[3]
 RN SEQUENCE FROM N.A.
 RP STRAIN-J138;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CML029 from USA."
 RL Nucleic Acids Res. 28:2311-2314 (2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39."
 RL Nucleic Acids Res. 28:1397-1406 (2000).
 DR EMBL: AE001628; AAD18591.1;
 DR EMBL: AJ133034; CAB37069.1;
 DR EMBL: AP002546; BAA98655.1;
 DR EMBL: AE002192; AAF38163.1;
 DR TIGR: CP0306;
 KW Signal;
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 928 OUTER MEMBRANE PROTEIN.
 SQ SEQUENCE 928 AA; 98332 MW; 58910A8F04F12219 CRC64;

Query_Match 11.5%; Score 585.5; DB 2; Length 928;
 Best Local Similarity 24.9%; Pred. No. 11e-31;
 Matches 266; Conservative 139; Mismatches 380; Indels 285; Gaps 46;

OY 1 MKKAEFFFLIGNSL-----SGLAREV-----PSRIFLMPNSVDP--TKESLSNKIS 45
 DB 1 MKSSLHWFLLISSIALPLSLNFSAFRAVEINLGPITNSGCGTTPPAQTINADCTIYN 60
 OY 46 LTGDTHNLTN-----CYLD---NLRYT---LAIQKTPNEGAAVTID----- 82
 DB 61 LTGDV-SITNAGSPALTASCCKETGNLSFGHGQYQLQNDAGANCTFTNTANKLL 119
 OY 83 -----YLSFPTQKEGIFFAKNLPESGCAIGVSPNPTVEIRITGPVFNNTCCR 136
 DB 120 SFGSGSYLSLIOT-----TNATTTGTA-----INSTGACSTQSNVSCYF 158
 OY 137 PFTSSNPNAAVKIREGAIHAQNLVYHNDVGFKNFSYVVGGAISTANTFVYSEHQ 196
 DB 159 GGNFNSMDN-----GGLQSSSISLSLNPLT--FAKKKATOKGALXSTGITTNNLT 209
 OY 197 SCFLFMONICITNTAGCGAIYAGTSNFSFNCDLFTINNACCA-----GGAISPTCS 252
 DB 210 NSASFSSEN-----TAANNCGAIYTEAS--SFISNKIAISFINNSVATSAATGAIY--CS 260
 OY 253 LTGNNGNIVFYNNRCFKNETASSEASDCGAIKVTIRLDTVNTNRGIFESDNTIKNYGGA 312
 DB 261 STS-----APKPYLLIS-----DNGNLNIGNTAITSGGA 290
 OY 313 IYAPVVLVNDGPTFYFINNIANNK-----GGAITYIDGTSNKSISADPHAFIENENTVNT 368
 DB 291 IYTDNLVSSGGPTLFKNNSAIDTAAPLGAIALDSGLSLSALGGDITFEENTVYVKA 350
 OY 369 NANGSTSNAPRRNAITYASSGGLL-LGASSONLLIYDPELV-NAGVSYSEFKED 426
 DB 351 SSSQITTT-----RNSINIGNTNAKIVQIRASOGNTIYDPTITISITALSALMLNCP 404
 OY 427 Q-----TGSVVFSGATYNSADPHOR-NLOTKTPAPLTLNSGFLICEHQAOLTVRFTQ 478
 DB 405 DLACGNPAYOGTIVTFSGEKISEAEADNLKSTIOQPLTLAGGQSLAKSGVYLVAKSFSQ 464
 OY 479 TGGVYSLGNLAVLSCYKKNAGNSASNA-SITLKHIGLNLSTILKSGAEIPLLWVEPTNNS 537

DB 465 SPGSTLLMD-----AGTTLETADGITTNNVLVNDSL-----KETKKA 502
 OY 538 NNVTADTAATFESLDVKKSLIDDYG-----NSPYESTDLTHALSSQPMISEASDN 589
 DB 503 TLKATQASQTVTLSC-SLSTVDPGSGNVYEDVSMNNQVYFSCLTTLADDPANHITDLAAD 561
 OY 590 QLRSDMDFSGLNVPHYXGQGLMTWCMATQDPEPARSSATITIDPOKANFHHTLLTLNLP 649
 DB 562 PLEK-----NPIHWGQGNMALSW---QEDTATKSKAAAT-----LTWTK 597
 OY 650 AGVYSPKHSPLANTLILGNML-----LATESLKNSEALTPSDHPWCITGGGLGM 702
 DB 598 TQYNPNPERGRLVANTLMGSEVDVRSIQVLATVATKROQETR-----GIMCGISNF 650
 OY 703 VYQDPRENHGFHMRSRSGYSAGMIAGQHT-----FSLKFSQTYTKLNERYA-KNNVSS 755
 DB 651 FPKDSKTKINKGFRHISAGY-----VVCATTTLASDNLITAAFCQLPGKDRDHFINKNRASA 706
 OY 756 KNYSGOGEMLFSLQGFLLTKLVG-----LSTYGDHNCHEHTYO--GEN-- 797
 DB 707 YAASLHLQHLATLSSPSLURLPGSESEQPVLPDAQISYISKNTMKTYTQAPRGESSW 766
 OY 798 -----LTSQGTFRSOTWGAVFDPDLPKPGSHIILTAPELALGIYS 840
 DB 767 YNDGCALELASSPLHTALSHGELFHA-----YF-----PFIKVEASTI 804
 OY 841 SLHSPTENG-AVPRSPSTKTLINVLVPIGVKGSFPMATQRPQAVTELAVQPVLYROEP 899
 DB 805 HODSFKERNTLVRSFDS-GDLINVSVPICITFEERSRNER-ASYEATVIYADVYRNKP 862
 OY 900 GIATOLLASKGIWFGSGSPSSRHA-----MSYKISQOQPLSMTLTIFQ 943
 DB 863 DCTTALLINNMTKTTGNLSRQAGIGRAGIYAFSPNLEVTNLSMEIR 912

RESULT 24
 ID 092895 PRELIMINARY: PRT: 978 AA.
 AC 092895;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN H FAMILY.
 GN PMP-14.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
 OX NCBI_TaxID=83358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CML029;
 RX MEDLINE=9920606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
 RL Nat. Genet. 21:385-389 (1999).
 DR EMBL: AE001628; AAD18596.1;
 DR INTERPRO: IPR000515;
 DR PROSITE: PS00402; BPD-TRANSP_INN_MEMBR; UNKNOWN_1.
 SQ SEQUENCE 978 AA; 103655 MW; 12D6610CBEB0871 CRC64;

Query_Match 11.5%; Score 583; DB 2; Length 978;
 Best Local Similarity 26.1%; Pred. No. 1.8e-31;
 Matches 263; Conservative 137; Mismatches 411; Indels 196; Gaps 45;
 OY 41 SNKISLTGTHNL--TNCYL-DNLRYIILAILOKTPNEGAANTIDYLSFFDTQKEGIYFA 97
 DB 67 SNNLTLKGLSLTFTSCGAPTNSNVAL-----SAAETLT-FKNF-----SSINFT 112
 OY 98 KNLTPESGCAIGVAPNSPFTVEIRDTIGVIFENN-----TCRPTSSNPNAAVKIREG 153
 DB 113 GNSGTGLGGLI-YGRD-----IVFOSIKDLIFTIRVAVSPASVTSATP--AITTVTTG 164

QY 798 LTSOGTFPSOTWGAFFDLP--MKPGSTHILTPFLGALGIYSLSHFTVEGAYPRSF 855
DB 765 KQSMQNNMFMNMLGASHSYPEYLHCFDT----YAPIYKLNLTIRRODSFSEKGEGRSF 820
QY 856 STKTPLINLVPIGVK-GSFMNATORPOAMVELAYQPVLYROEPGIATQLLASKGIWFG 914
DB 821 DQSN-LFNLSTLPIGVKFEKFSDCND--FSYDLTSLVDPDLINRDKCTALVLSGASWET 877
QY 915 SGSPSSRHMSYK 927
DB 878 YANNLAROLVOYR 890

RESULT 21
Q9RB65 PRELIMINARY: PRT: 928 AA.
AC Q9RB65: Q9RB64: Q96P2: 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (OUTER MEMBRANE PROTEIN 5
DE PRECURSOR (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN PMP_10 OR OMP5 OR CP0303.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V1310;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CRL029 from USA."
RL Nucleic Acids Res. 28:2311-2314 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-V1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGAI-repeat belong to a subfamily
RT of autoexporting pathogenicity factors."
RT Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE-20150255: PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umeyam L.A., Ulteback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J.,
RA McManary G., Salzberg S.L., Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406 (2000).
DR EMBL: AP002546: BAA98657.1;
DR EMBL: AJ133034: CAB37071.1;
DR EMBL: AE002192: AAF38160.1;
DR TIGR: CP0303;
KW Signal.
SQ SEQUENCE 928 AA: 97229 MW: 0590D5206A1D0E1 CRC64;

Query Match 11.5% Score 586.5 DB 2: Length 928:
Best Local Similarity 25.6% Pred. No. 9.6e-32;
Matches 264; Conservative 143; Mismatches 377; Indels 249; Gaps 48;

QY 1 MKKAFFFLIGNSL-----SGLAREV-PSRIF-----LMPNSVDPPTKESLS 41
DB 1 MKSQSMVLVSLTACFTSCSTVFATLKNIGPSDFSDSTNGTTPKNTTGTGIDYTLT 60
QY 42 NKISL--TGDTNLT-NCYLDNLK-----YIALLIO-KTPNRCGAATIT----- 81

DB 61 GDITLQNLGDSALTKGCFSDTTESLSFAGKQYSLSPFLNIKSSAGCALSVTTDKNLST 120
QY 82 --DYLSFFDQKEGIYFAKKNLPRESGALIGVAPNSPVEIRDTIGPIFENNTCCRFPT 139
DB 121 GFSSLTFLAABSVI-----TTPSGKAVKCGDUL-----FDNNGTILFKODIC----- 165
QY 140 SSNPAAVANKIREGALHAONLYTHNNDVYGFNMKFSYVRGAIISTANTFVSENOSCF 199
DB 166 -----EENGCAISTKNLSKNTSGSISEGKNS----- 193
QY 200 LFMNICTIOTNAGKGAIIYAGTSNFSFNCDLEFINN-ACCAGCAIFSP-ICSLTGNR 257
DB 194 -----SATGKKGGAICATGIVDITNTPTLFSNNIAEAGCAINSTONCTITGNT 244
QY 258 GNIVEYNNRCFKNVETSEASDGAIVKVTTLDTVGRNGRFFESDNTTKNGCAIVAPV 317
DB 245 -SLVSESN-----SVTATGNGGALSGDADVTISGQOS-VTSGNOAVANGCAIVAKK 295
QY 318 VTLVD--NGPTVFINNIAN---NKGAIYIDGTSNKRISADRAHIIENENIVVTNAN 371
DB 296 LTLASGGGGGGSFNSNNIYOGTTAGNGAISIILAECSLSAEGDITNGNAI----- 348
QY 372 GTSTAMPNRRNAITVASSSGEILLGAGSSONLIFDPIEVSNAGV---VSEFK----- 423
DB 349 -VAITPOTTKRNSIDIGSTAKITNLRAISGHSIFFYDPTTANTADSTDTLNLKADAGN 407
QY 424 EADDTGSYVESGATVNS-----ADFHORNLOTKTAPLILNSGFLICIDHNOIVNRTQ 478
DB 408 STDVSGSIVFSGEKLSEDEKAYAD---NLVSTLKOPYTLTNGNLVLRGTLDTKGTQ 463
QY 479 T-GGVVSLGNGVSLSCYKNGAGSNASNTLKHIGLNSITLKSAGBIPLIMVEPTNNS 537
DB 464 TAGSSVINDAGTTL-----KASTEVELTIGLSIFVDS-LGSGKKVYI----- 504
QY 538 NNYTADTATPSLSYKUS---LIDDYGNSTPESTD--THALSSQPMLSISEASDNL 591
DB 505 -----AASASAKNVALLSCPTILLDNOGNA-YENHDLGKTODFSFVOLSALGTATTVDV 556
QY 592 RSDMDDFGSLNVP-----HYGMOGLMTGMNAKTQDPEPASSATITDQKNRFRITLL 645
DB 557 PA-----VPYATPPIHYGQGTWGTWVDVDTASIPKT-----KIATL 593
QY 646 TWLPAGVYSPKRRSLPLANTLGMNM--LATES-LKNSAELTPSDHPFGITCGGLGM 702
DB 594 AMTNGTGLNPNRPGRLPNSLWMSGFSFDAQIGYIERSALLTLCDRGF--AAGVANF 650
QY 703 VYODPRENHPGFHMRSSGYSAGMIAGOT---HTFSLKFSQYTKLNERAKNNVSKYNS 759
DB 651 LDKDKRGEKRYRHKSGYALOG-AAQCSEMIISFAFCQLGFS-----DKDFLVAKNHT 704
QY 760 CGCEMLFSLQE-----GFLITKLVG-----LYSGDINCHIFTY-QDEN 797
DB 705 DTYACAFYIQHTIECSGFLIGCLLDKLPQSWSHKPLVEGQLAVSHVSDLLKTKTYAYREV 764
QY 798 LTSOGTFPSOTWGAFFDLP--MKPGSTHILTPFLGALGIYSLSHFTVEGAYPRSF 855
DB 765 KQSMQNNMFMNMLGASHSYPEYLHCFDT----YAPIYKLNLTIRRODSFSEKGEGRSF 820
QY 856 STKTPLINLVPIGVK-GSFMNATORPOAMVELAYQPVLYROEPGIATQLLASKGIWFG 914
DB 821 DQSN-LFNLSTLPIGVKFEKFSDCND--FSYDLTSLVDPDLINRDKCTALVLSGASWET 877
QY 915 SGSPSSRHMSYK 927
DB 878 YANNLAROLVOYR 890

RESULT 22
Q92813
ID Q92813 PRELIMINARY: PRT: 947 AA.
AC Q92813: 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

084880
ID 084880 PRELIMINARY: PRT: 1016 AA.
AC 084880:
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE PUTATIVE OUTER MEMBRANE PROTEIN H.
CN PMPH.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX:
RC MEDLINE=99000809; PubMed=9784136;
RX Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.:
RA "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
DR EMBL: AE001360; AAC8470.1;
DR INTERPRO: IPR002016;
DR PROSITE: PS00435; PEROXIDASE_1; UNKNOWN_1;
SQ SEQUENCE 1016 AA; 107904 MW; E691912C3A2BDE67 CRC64;

Query Match 11.7%: Score 595; DB 2; Length 1016;
Best Local Similarity 25.4%; Pred. No. 2.9e-32;
Matches 261; Conservative 149; Mismatches 404; Indels 214; Gaps 49;

QY 42 NKISLTGDTHT-----NCYLONLRYLILQKTPNEGAAYITDY, SFFDTQKEGIYF 96
DB 69 DMLTTIGONHTLSTFDSCGPVLONTAFISA-----GETLTKPFSS-----LMF 112
QY 97 AKNLTPESGALCYASPSPFVEIRDTIGVIFENNT-----CCRP 137
DB 113 SKNVSGCEKGM-----SGKTYSISGA-GEVJFMDNSVGYSLIVPASTPTPPAPAPAP 166
QY 138 FTSSNPNAVNKIREGGAHAQNLINHHNDVYGFKNSTYVAGAISTANTVSENO5 197
DB 167 AASSSLSPYSDARKSIFSVET-----SLGISVKKGVNF-----DNNAG 207
QY 198 CF--LFMDNLCIOTNTAGKAGIYAGTSFESNCD--LFFINNACAGAIFFPICS 252
DB 208 NFGTVRGN--SNNNGSGS-GSATPTSFYKCKGKXSFIDNVA5CGGVYKGTVL 263
QY 253 LTRGNRNIYFYNRRCKNMTTASSEASD-----GGAI-KVYTRLDVTGNRGRIFSD 303
DB 264 FKNNEGIFFRGNTAYDLDLILATSHDQNTETGCGGCVICSPD5VKFEGNKSIVFDY 323
QY 304 NTRKNTGGAIAVAVTLVDNGPTFYFININNGGAIY--IDGTS-----KISAD 353
DB 324 NFAKRGCGSLTRKESLVADDSVAFSNNTEKKGAIYAPTDISTGSGILTEBRNRAE 383
QY 354 RHAIIFNE-----NIVVTNTA--NCTSTANPPRRNATIVASSGEBILGAGSSQ 402
DB 384 GGAICVSEASGSGNTLTLSDDDIYFSGNMTSDRGERSAARILDDGTIVSLNAGSL5 443
QY 444 KLIFYPDVYVNN5AAGASTPSSSSMPGAVTINOSGN--GSVIFLTAESLTPSEKLYVLN 501
DB 448 LOTTPAPLISNGFLIEDHAQLTVRPTQGTGVSILGNGAVLCYKNGANGASNASI 507
QY 502 STSPNPALTYSGGELVYVTEGATITGTITATSGRVTLG5ASISAVAGAANN--NTC 558
DB 508 TLKHTGLNLSILKSGAETPLIWEPTNNSNNYADTAATESL5DKVILSIDYGN5PYE 567
QY 559 TVSKLIGIDLEFLTPNKITAILGADG-----VTVNSGSTLDL--VMESEAEVYDN-PLF 610
DB 568 STDLTHALLSOPMISIS5AS-DNQLRSDMDPFCGLIAPHYMGGLMTWGMKATQDDPEAS 626

DB 611 VGSILT-----IPVTLSSSSASNGVTKNSVTINDADAAHYCGGWSADMTK---PPLAP 662
QY 627 SATITDPOKANRFRHTLLTLWLPAG-----YVSPKHSR5PLIANTLW--GNMLLA-TESLK 679
DB 663 DAKGVPPENTN---NTLTLWRPASNYGEYRLDPQRK6ELVPNSIMVAGSALRTFTNGLK 719
QY 660 ----NSAELTPSDHPHFGITGGCIGMMVYODPRENHGCFHMRSSGYSAGMIACOTH--- 731
DB 720 EHYVSRDVGFAVSLI-----ALGDYILNVTODDRFGFLARFGQA--TASHVEN 768
QY 732 --TESLKFSQTY--TKLNBRYAKNNVSSKNYSCGEMFL5LOEGFLTLKVLGY--SYGD 785
DB 769 GSIFGVAFGOLYGOTSKRNY5SKDAGNMTLSCFGRSVVDIKG---TETVWYMETAG- 823
QY 786 HNCHEIFTYOGENTL5QGTFR5Q-----TMGA---VEFDLPMKPF5THILTA-- 830
DB 824 YSVHMHNTQYFNDKTQKRDH5KCHWNNNYAFVGAENHNFLEYCIPTROFARDYELTGFM 883
QY 831 PFLGALGIYSSLSHFT5VGAATPR5STKTPILNVLVPIGVK5CFMNATQ--POAWTYEL 888
DB 884 RFEMAGWSSST--RETGSLTRYEFAR5G--HMSLPIQIYAAVASHVR5SPSKLTLM 939
QY 889 AYOPVLYRDEPIATOLLASKGIMFGSGSP5SRHMSYKISQOTOP5LWTLHPQYHGFY 948
DB 940 GYRPDIWKVTPHCHMEIITANGVKTPIQ55PLARHAF5LEVHD-----TLXIHHC-- 989
QY 949 S5STFCNY 956
DB 990 --RAYMNY 995

RESULT 19
QJ9J52
ID 09J52 PRELIMINARY: PRT: 947 AA.
AC 09J52:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE POLYMORPHIC MEMBRANE PROTEIN A FAMILY.
GN PMP-19.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RC MEDLINE-20330349; PubMed-10871362;
RX Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishi K., Hattori M., Kuhara S., Nakazawa T.:
RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CwL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AP002547; BAA98745.1;
SQ SEQUENCE 947 AA; 103628 MW; 9CBPDAF290A771EC CRC64;

Query Match 11.5%: Score 587; DB 2; Length 947;
Best Local Similarity 25.2%; Pred. No. 9.2e-32;
Matches 250; Conservative 174; Mismatches 389; Indels 178; Gaps 47;
QY 14 LSGIAREVSPRIPIIMP5VDPDKESISNNK--SLTGDTHNLTNCLNRYLIALIOLK 70
DB 30 LSGHSEDELEFLTRSS5PTKTIYSLRKDFIYCDFGNSIHKRGAAFLLWKGDLEFINS 89
QY 71 TPNEGAAVTITDYSFDTQKEGIYFAKNLTPESGAIGV5APNSPVE-----IRDT 123
DB 90 TPL--AAITTKNI--HIGARGAGL5SESNT-----FKGL5LVLNNESMGVLT 137
QY 124 IGVYIFENNT--CCRP5TPNPAVNKIREGGAIIHON-----LYTNHNDVYGFKNF 176
DB 138 SGDL5FINNTSVLCQNNISYGP-----GGALLQGRKSKALFFRDNGTILFLKNK 188
QY 177 SYVR-----GCAIS7ANFV5SENG5CFLFMN--ICIOITNAGGAIYACT--SNSF 226

```

OY 104 SGAIGYASPNSTVEIRD-----IGPVIFENNTCCRPFTSSPNNAVKIREGAIHMO 159
DB 103 AOSNMAAATAADKADKLTPTGFSNLSFIAAPGTTVASGSLSSAGALINTDNGTLTFFSQ 162
OY 160 NLYINNHVDVGMKNFYSVGAIASTANTFVSENOGFLFMDNICIOTNTAGK-GGAI 218
DB 163 NYSNENNN-----GGAIIT-KTLSISGNTS-----STPTFSNNAKKIGGAI 203
OY 219 YACTSNFESNNCDLFFINN-ACCACGAI-FSPICSLTGNRNIYVYNNRCEKNVETASS 276
DB 204 YSSAASISGNTGOLFEMNNKGTGGALGFEMASSITON-SSLFESGN-----TAD 255
OY 277 EASDGAIV-----TTRLDVGNRRIFFESDITKNYGAIAVPVTVVDNGPTVEYINN 331
DB 256 AAGKGAIYCEKGTETPTLTISGNS-LTFAENSSVYOGAICAHGLDIASAAPTLPFSA 314
OY 332 IANN-----KGAIYIDGTSNSKISADRAHIIENENIVTNTANSTSTSNPRRAITV 387
DB 315 RCGNTAAGGGAIAIADSGSLSSANOGDITLGNLT-----TSTAPSTRNAIYL 366
OY 388 ASSSGEITLGAASSONLIFYDPIEVSNAGVS-----VSFNKADQTSVPSGATVNS 440
DB 367 GSSAKITNRAAOGOSIYFDPASNTGASDYLITNOPDSNPLDYSIGTIVSEKLS-S 425
OY 441 ADEHOR--NLQTKTPAPLTLNGLFCIEDHQAULTVNRFTQOTGVSVSLGNVLSCKYKNGA 498
DB 426 ADEKADADNTSLKQPLASGTLAKGNVELDVGTFOTEGSTLL-----NQPGT 477
OY 499 GNSASNASTITLHIGLNLISLKSAGIPLWEPNNSNNTADTAATESLSDVKLSLI 558
DB 478 KXKADTEALSLTKLVYDLDAL-----EGNKSYSIETAGANKITTLTS-PLVQ 524
OY 559 DDYGNSPYSTDLTHALSSQPMLSISEASDNQLRSDMDPSGLNV-----PHYMOGL 611
DB 525 DSSGNF-YESHITNQAF-TQPLVVFYAT--AASDIYDALTLSPVOTPREPHYGOCH 578
OY 612 WTGWMKKTDDPEPASATITDPOKANRPHRTLLTLPLPGYVSPKHSPLINTITMGNN 671
DB 579 WEATMA--DTSTAKSGT-----MTWVTGYNDNPERKASVVDSDSLWASP 620
OY 672 LATESLKSNAELTPSDHPFW---GITGGGLGMVYODPRENHPGHRSSGYSAGMIAG 728
DB 621 --TDRTIQIQTISQANSIYQORGLMAGTANFHKDKSGTQARRHKSUYGIVGSAE 677
OY 729 --OTHTFSLKFSOTYTKLNERIAKNNVSKNYSQ-----GEMLPFSLQ 769
DB 678 DESENIFSAFCOLFCKDKDLFIVEN-TSINYLASLYLQHRAFLGLGMPSPFSITDMK 736
OY 770 EGFLLKLKLVGLSYGDNCHNFT-----OGENLTSOGFRSOTMGGAFFDPMK-PTG 823
DB 737 DPLILNAOLSTSYTKNDMDTRITSTPEAOGSVTNNSGALE---LGSIALYLPKRAPPF 793
OY 824 STHILAPLGAIGIYSLSHFTVEGAYPRSESTKPLINVLVPIGVKGSFMMATORPOA 883
DB 794 OGTF--PFLKQAYVSRQONFESGAEARAFD-DGDIVNCSIPVIGIRLEKISDEDK-NN 848
OY 884 WTVELAYQVLYRQEPGIAITOLIAKGIW 912
DB 849 FEISLAVYIDVYRKPNRSRTSLWVSGAW 877

```

RESULT 16

```

OY 086164 PRELIMINARY: PRT: 928 AA.
DB 086164:
AC 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE OUTER MEMBRANE PROTEIN 4 PRECURSOR.
GN OMPA OR PMP_11.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

```

```

OX NCBI_TaxID=83558:
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CDC/CWL-029/YR-1310;
RA Knudsen K., Madsen A.S., Mygind P., Christiansen G., Birkelund S.;
RT "Identification of two novel genes encoding outer membrane complex
associated surface layer proteins in Chlamydia pneumoniae."
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=VR1310;
RA Hjerno K., Boesen T., Daugaard L., Knudsen K., Madsen A.,
RA Christiansen G., Birkelund S.;
RT "Chlamydia proteins containing the GGA1-repeat belong to a subfamily
of autotransporting pathogenicity factors."
RL Submitted (Jan-1999) to the EMBL/Genbank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RA MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
RL Nat. Genet. 21:385-389(1999).
RN (4)
RP SEQUENCE FROM N.A.
RC STRAIN=CJ38;
RA MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA."
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL: AJ001311; CAB04672.1;
DR EMBL: AJ133034; CAB37072.1;
DR EMBL: AE001628; AAD18593.1;
DR EMBL: AP002546; BAA98658.1;
KW Signal.
FT SIGNAL 1 17
FT CHAIN 18 928 POTENTIAL.
FT CHAIN 18 928 OUTER MEMBRANE PROTEIN 4.
SQ SEQUENCE 928 AA; 98903 MW; 788BCDD62C911402 CRC64;

```

Query Match 11.7% Score 597; DB 2; Length 928;

Best Local Similarity 26.1%; Pred. No. 1.8e-32;

Matches 253; Conservative 148; Mismatches 361; Indels 206; Gaps 47;

```

OY 36 TRESLSNKSISLTDG-----THNLTCY--LDNLRY-----DAILOTPEGA 76
DB 45 TPKTSATITSLTGCDVFEYEPKGTPLSDSCFQKOTNTMLFLGHSLTGCFIDAGTHAGA 104
OY 77 AVITIDYLSFPTQKEGIYFAKNLTPESGAIQYASPNPVEIRDTIGPVIFENNTCCR 136
DB 105 AASTT-----ANKNLTFSGFSLSFDSPTTV---TTGQ----- 136
OY 137 -----GTLSSAGGVNLENI---RKLVAAG---NESTADGAIIGA-SFLTGTGS 178
DB 197 SCFLFMDNICIOTNTAGKGAIAVAGTSNFSFNCDL--FFINNACCAGAGFSPICSLT 254
DB 179 GDALEFSNN-----SSSTKGAI-ATTAGARIANNITGYVFLSLIASTSGAIDDETSIL 232
OY 255 GNRGNIVFYNNRCFKNVEFASSEASDGAIKVTRLD---VTGNRRIFFSDNITKNYG 310
DB 233 SNKKFLYFEGN-----AAKITGAIICNTKASGPELLISNNKTLIFASNVAETSG 282
OY 311 GAIYAVVTVLVNDGPTFYI-NNI--ANNKGAIYIDGTSNSKISADRAHIIENENIVTV 367
DB 283 GAIHAKKIALSSGCTFEFLNNVSSATPKGAIISDASELSISATGCTITVRRNTLT-- 340
OY 368 TNANGTSTSNPRRAITVASSSGEITLGAASSONLIFYDPI--EVSAGVSVSEFNKEA 425

```

```

Db 519 TCKEYENHKLNDTLA--LGGIOLG-----GKSVTTTNVSHV----- 555
Oy 591 LRSDMDFSGLVNPHYGMGLWTMGWATODEPASPATITDPOKANRPHRTLLTLPLA 650
Db 556 -----GVAETHYGYGNGNSVSMVKDNNSDPKT-----QTAIFTMNKT 592
Oy 651 GVPSPKRRSPLIANTLMGNMLLATESLKNSAELTPSD--HPPWGITGGGLGMVYODRR 708
Db 593 GVPNPERAPLVLNLSMGS-FIDLRISQDVLERSVDSILETRGLMVGSGNFFHKDRN 651
Oy 709 EHHPPGHMSSSGYSAGMINGOth--TFSLKFSQTYTKLBERAKANNVSS----- 755
Db 652 AENRFRHISISGYVGLATTNTSREDLSVAFQDLPAKDDYLVSKAANVAVGASYOYH 711
Oy 756 -----KNYSCGEMFLSLOEGFLTKLVLYSGVDNCHHFTGEGNLTSGT 803
Db 712 SKFDLITLRFNPNPTCCSG---FSKEIRIFLDQAQI--TYCHTANNMTSTYDPEV--KGS 765
Oy 804 FRSQMGAVFEDLPKPKPGSTHILT--APFLGALGYSSLSHTEVAVPRSTKPLI 862
Db 766 WONDITGLTSLTSPRIPVSSSIFDSYAPFAKQYVYAHODDFKEPTTEGRAVES--SDLL 824
Oy 863 NVLVPVGVGSMNATQROAMTVELAYQVLYROEPGIATOLLASKGIMFGSGSSRH 922
Db 825 NVSPVIGIKFEKLYGER-SAYDLTLMYIPDYVRHNPSCMTGLAINDVSMLTATLNAQ 883
Oy 923 AMSYKISQOTPLSMTLTHFOYHGFSSSTFCVY---LNGEIA 962
Db 884 AIVRAGNHIALTSGVEMFSQF-GFELRSSSRNYNVDLGAHY 925

RESULT 13
O9P145 PRELIMINARY: PRT: 987 AA.
ID 09P145:
AC 09P145: 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE POLYMORPHIC MEMBRANE PROTEIN G FAMILY.
GN TC0263.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
  White O., Hickey E.K., Peterson J., Uterback T., Berry K.,
  Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
  Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
  Eisen J., Fraser C.M.;
  *Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
  pneumoniae AR39";
RT Nucleic Acids Res. 28:1397-1406(2000).
RL 12
RN SEQUENCE FROM N.A.
RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
  Hickey E.K., Peterson J., Umayam L.A., Uterback T., Berry K.,
  Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
  Dodson R., Gwinn M., Nelson W., Deboy R., Kolonay J., McClarty G.,
  Salzberg S.L., Eisen J., Fraser C.M.;
  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEO02293; AAF39132.1;
  TIGR: TC0263;
  DR SEQUENCE 987 AA; 104867 MW; 320798D6BEB2DA42 CRC64;
  SO

```

```

Db 82 NLGNFTTAGRSHSLV---FENIR-----TSTNGALSNHAPSGLVIEAPDELSTL 130
Oy 88 DTQKEGIVFAKULTPESGGAIGVAPSNPTV-----EIRDTIGPIFENNTCCRPST 141
Db 131 NCNS-----LVSVFOTGTT--TSVPSNCTIYSRDVLURD--IKKVSFSLVS----- 177
Oy 142 NPNAAVNKIREGAIHAONLYINHNDVYFMKNFSYVBAKISTANTFVSENOGCFLE 201
Db 178 -----GDCGAI DAOGLMNVNGIEKICTFOENVAQSGCAGCYTKTPSAVGNVPLSF 228
Oy 202 MDNICIQTPAAGKAIYA-----GTSNPF-----SNCCDLFLINNACCAGALFSPIC 251
Db 229 LGNV-----AGNKGQGVAAVKDQGGAGATDLSVNEFANTVAVEFENSRARIGGIYSD-- 281
Oy 252 SLTGNRGIVFVNNR---CFKNVETA-----SSEASDGAII--KVTTRLD 291
Db 282 -----GNISFLGNKATVFLSNVAPLIYVDPAAAGQPRADKDNVGDGALFCKNDNTIG 335
Oy 292 VTG--NRGRIFESDNTIKXNGALYAVVTLVNDGPTPIINNANNGGAIYIDGTSNK 349
Db 336 EVSFEKQGVVFFSKNIAAGKGAIVAKKLTISDCGPVQFLGVAND--GCAIYLVQDGLS 394
Oy 350 ISADRAITFENIYVNTNANCTSTANPPRRNATTVASSGEILLGSGSSONLIFYP 409
Db 395 LSADRGDIIFDGNLKRMAIQAATYADV--MVASNAISMATGQIITTLRAKEGRRLIFNDP 453
Oy 410 IEVSNAG---VSVPKQADQTSVVS--GATVNSADFHQRNLQTKTPAPLTLNSGFLCI 465
Db 454 IEMANGQPIQITLVNEGCGYTGDIYFAKGDV-----LYSIELSGRITL 500
Oy 466 EDHAQLTVNRFTQGVVSLGNGAVLSVCYKNGAGNSASITLKHIGLNSILK--SGA 524
Db 501 REQTKLLVNSLTQGGVHMEGSLDF---AVTPPANSMAITVHESLASLTKNNGV 557
Oy 525 EIPILNVEPTN-----NSNNYTAOTAA--TFSLSDVLA,SLIDYGNPSTLDTLH,SSQ 578
Db 558 TNP-----PTNPVYVSSAVYIGNVAGVTTISGP--IFFEDIDETAYANNOMLGADOTI 610
Oy 579 PMLISEASDNOLRSDDPGSGNLNVPYHGMGLWTMGWATODEPASPATITDPOKANR 638
Db 611 DVLQHLGANGPRANATDTLIGNESSKYQGGSWITQW---EPDA-----NPPQNN 660
Oy 639 FHRITLLTLMIPAGYVSPKRRSPLIANTLMGNMLLATESLKNSAELTPSDHPFW-GITGG 697
Db 661 Y--MLKASMTKTGYNGBRPAVSLVNSLMGS--ILDVRSASHAIQASIDGRAVRCGIWIS 717
Oy 698 GIGMAYYODPRENHPGFHMRSSYSAGMIA--GQTHFSLKFSQTYTKLBERAKANNV-- 753
Db 718 G1SNFFYHODALGQGYRHISGYSIGANSYFSS--MFGIATFETIGR-----SKOYVVC 771
Oy 754 SSKNYSGCEMFLSLOEGFLTKLVG-----LYSGDNCHHFTYOGENTLSQGTFRSQ 807
Db 772 RSNDDHTCVGSVYLTSQLGSCLEFCDATVRASYGCGNCHMKSYFADE--SNVMDNN 829
Oy 808 TMGCAVFPDLPKPKPGSTHILT--APFLGALGYSSLSHTEVAVPRSTKPLINLV 865
Db 830 CVYGEVAGAGLPIMLASKLYLNLRFVQAEFAYAEHSESTERGDOARFEKS--GHLMN 888
Oy 866 VPIGVKGSFNANATQROAMTVELAYQVLYROEPGIATOLLASKGIMFGSGSSSHANS 925
Db 889 IPGVGFVD--RCSSKHPNKSIFMGAYICDAVRSIGETITLLSHKLTWTDAFHLAHGV 947
Oy 926 YKISQOTPLSMTLTHFOYHG 946
Db 948 VRGSMYAS---LTGNIEVYG 964

RESULT 14
O9RB66 PRELIMINARY: PRT: 930 AA.
ID 09RB66:
AC 09RB66: 01-MAY-2000 (TREMblrel. 13, Created)
DT

```

```

Query Match 12.2% Score 620.5; DB 2; Length 987;
Best Local Similarity 25.8% Pred. No. 5e-34;
Matches 253; Conservative 153; Mismatches 404; Indels 171; Gaps 42;
Oy 39 SISNKSISLTGDTNLTNCTYLDNLTALITLAIQKTPNEGAA-----VTITDYLSSF 87

```

```

Db 713 LMGS-IIDIRSAHSAIQSVDCRSYCRGLAVSGVSNFEYHDRDALCGGYRYISGYSIG- 770
OY 726 IAGOTHTFLKFSOTYTKLNERYAKNNV--SSKNVSCGEMFLSIOEGFLTLKVLGYSY 783
Db 771 --ANSYFSSSMFGLAFTEVFGFR-SKDYVVCNSNHACIGSYVLSTKQ-----LCGSLF 822
OY 784 GDHNCHEHYTGC-ENLTSGTFRSOT-----MGAVFEDLPMPKPGSTHILT--APF 832
Db 823 GDAFRASYGNGCHMKSTYFAESDVRMDNCLVGLIGLPIVITPSKYLINELRPF 882
OY 833 LGALGIYSSLHFTFVGAVPPSFSTKTPPLINVLVPIGVKGSFMATQRPQAVTVLAYOP 892
Db 883 VQAFPSYADHSEFTFEGDQARAFRS--GLMLNLSYVGVKFDRCSSST--HPNKYSFMGAYIC 940
OY 893 VLYRDEPGIATOLLASKGIWFGSSSPSRHAMSYKISQOTOPLSMLTLHFQYHG 946
Db 941 DAYRTISGQTTLTSHOETWTTDAFHLARHGVIRGSMYAS---LTSNIEVYG 990

RESULT 10
O92898 PRELIMINARY: PRT: 936 AA.
ID 092898
AC 092898:
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY.
GN PMP-7.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN NCBI_TaxID=83558;
RP SEQUENCE FROM N.A.
RC STRAIN-CWL029;
RX MEDLINE-920606; PubMed-1019238;
RA Kalman S., Mitchell W., Marathe R., Jammal C., Fan J., Hyman R.W.,
RA Olinger L., Gitchell J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 23:385-389(1999).
DR EMBL: A6001627; AADI8589.1;
SQ SEQUENCE 936 AA: 100079 MW: 888107853D194EC CAC64:

```

```

Query Match 12.7% Score 645; DB 2; Length 936;
Best Local Similarity 27.0% Pred. No. 9.9e-36;
Matches 262; Conservative 146; Mismatches 392; Indels 172; Gaps 44;

```

```

OY 75 GAAVITTDYLSFDTQKECIYFAKNLTPBSGAFGY-----ASPNSPTVEIR 121
Db 54 GTTYSLSLSDVSFQNALGIPILASGCFLEKGGDLTFQGNHAKFAFINAGSAGTAVST 113
OY 122 DTGPIVIFENN-----TCRPFSSNPRAANRKIREGAIHAONLXHNHNDVVGFM 174
Db 114 SAADKMLFNFDFSRLLSIISCPULLSPGQALSKV-----GNLSLGNQSQIIFIQ 164
OY 175 NFSYVGGAISTANTFVSNOSCFLEMDNICIQTNTAGKAIYA-GTSNSESNNCDL 233
Db 165 NFSSDNGVINTKN-FLLSSTQSPASFSRN---QAFTRKKGCVVYAKATITENSPIVS 220
OY 234 FTINMACAGCAIFS-PICSLTGNRCNIVEYNNKCFKNVETASSEASDGAIKVTT--R 289
Db 221 FSNQLMAGSGALYSTNCISITDR-FQVIFDGSAMWAAO-----AQGALICCTTIDKT 273
OY 290 LDTVGNRGRIFESDNTKKNKGAIYAPVTVLVNPGPTFFINNIANKN-----GGAIIIDG 344
Db 274 VTLTGNK-NLSFTNNTALTYGALSGLKYSISAGCPPLTFQSNISGSSAGCGGGAIIIAS 332
OY 345 TSNKISANDRAIIFENENITVNTNANGTSTSANPRRAITVASSSGEILLGAGSSQNL 404
Db 333 AGELALATSQDITFNNQVTN-----GSTF-----RMAINIIDTKVYSIRAAATGOSI 382
OY 405 IFYDPIEVSNAQVS-----FNKEADQGTGSVVFSGATVNSAD-FHQRNLQTKTP 453

```

```

Db 383 YFYDPI--TNFGTAASDTLNLNLADANSELEYGAIVFSEKISPEPKALAAVNTSTIR 440
OY 454 APLTSLNGFLCIEDHAQILVNRFTQGVVSLGNGAVLSCYKNGAGNSASNASITLKHIG 513
Db 441 OPVILARDDLVRKGVVYTFEFDLQSPGSRILMDC-----CTTISAKFANLSLNGIA 492
OY 514 LNLSSILKSGAEIPLLVNEPPIVNSNNTADTAATFSLSDVLSLIDYDGNSPYSTDLTH 573
Db 493 VNLSL-----DGTNKALKTEADKNISLSGT-IALIDEG-SFEYENHLKS 538
OY 574 ALSSQPMLSSEASON-OLRSDMDPFGSLNP--HYGQGLMTGMWATQPEPPASSATI 630
Db 539 A-STYPLLELTAGANGTITGALSTLTLQEPETHYQGMWOLSMAN-----ATSKSI 591
OY 631 TDPOKANRFRHTLLTWLPAGVYVSPKURSPLIANTLWGNMLLATESLKSNAELTPSDHP 690
Db 592 GS-----INWTRTGYPSPERKSNLPLNSLWGN-FIDIRISINQLETKSSGEP 638
OY 691 F-----WGITGGGLGMVYODPRENHPGFHMSSGYSAGMIACQRTTFSLKFSQYTXLNE 746
Db 639 FERELM-----LSGIANFERYDSMPTRHGFRIHISGVALGITAATPAEDQLTA--FCQLFA 693
OY 747 RYAKNNVSSKN-----YSCQEMFLSLOEGFL-----LTKLYGL----- 780
Db 694 R-DRNHTITGKNHGDYVYGLFHHTEGLFDI-ANFLMKATRAPVLSLEISQIPLSPDA 751
OY 781 --YSYGDHNCHEHYTQGENLTSGTFRSQTGAVFEDLPMPKPGSTHILT--TAPFLGA 835
Db 752 KFSYLAHTDNHMKTYT--DNSILKGSWRNDARFADLCAIGSLF-VISVYVLLKEVPEPVKV 808
OY 836 LGIYSSLSHFTFVGAVPPSFSTKTPPLINVLVPIGVKGSFMATQRPQAVTVLAYAOPVL 894
Db 809 QYIVAHQODFEYRIHAEGRAF-N-KSELINVEIPIDGV--FERDSSSEKGTIDLTMLYLD 864
OY 895 -YRDEPGIATOLLASKGIWFGSSSPSRHAMSYKISQOTOPLSMLTLHFQYHGYSSTF 953
Db 865 AYRRNPKCOTSLIASDANMAYGTNLAROGFSVRAANHQVVPNHIEIQOF-AFEVRSS 923
OY 954 CNYLNGELALRF 965
Db 924 RNV-NINLGSKF 934

```

```

RESULT 11
O93542 PRELIMINARY: PRT: 936 AA.
ID 093542
AC 093542:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE POLYMORPHIC OUTER MEMBRANE PROTEIN G FAMILY (POLYMORPHIC MEMBRANE PROTEIN G FAMILY).
GN PMP-7 OR CP0308.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN NCBI_TaxID=83558;
RP SEQUENCE FROM N.A.
RC STRAIN-J138;
RX MEDLINE-20330349; PubMed-10871362;
RA Shitai M., Hatakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kihara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AR39;
RX MEDLINE-20150255; PubMed-10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,

```

QY 960 EIALRF 965
 Db 941 GTTFKF 946
 RESULT 7
 092880 PRELIMINARY: PRT: 946 AA.
 AC 092880: 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE POLYMORPHIC OUTER MEMBRANE PROTEIN E/F FAMILY.
 GN PMP-18
 OS Chlamydia pneumoniae (Chlamydia pneumoniae)
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia
 NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CML029:
 RX MEDLINE-99206606: PubMed-10192388;
 RA Kallan S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RI Nat. Genet. 21:385-389(1999).
 DR EMBL: AE001631: AAD18610.1: -;
 SQ SEQUENCE 946 AA: 103655 MW: C98B904E863C452C CRC64:

Query Match 18.6%; Score 948.5; DB 2: Length 946;
 Best Local Similarity 29.1%; Pred. No. 1.9e-56;
 Matches 281; Conservative 167; Mismatches 417; Indels 101; Gaps 34:
 QY 35 PTKESLSNKISLTGDTNLT-----NCLVDMLRYLALQTPNKGAVTTITDY 84
 Db 47 PLIDLTIN--MTPVSHRATLFGVRDDTODIYLDHONSIESWFENFSOGGALSCS-L 102
 QY 85 SFEDTQKEGIFYFAKNLTPESGAIYGAASPNPTEIRDTIGVIFENNTCCRFSTSNPN 144
 Db 103 AITNT-KNQLFLNSFAIKAGAM-YVNGN---FDLSENIIGTIFSGNL-----SFPN 150
 QY 145 AA--VVKIREGAIHAQNLTNNHNDVYCFMKNFVYRGCAISTANTFVYSENQCFLEM 202
 Db 151 ASNFADCTCGVALCSNVTISKNGAYFINKKAKSSGGAIOAIIINKDNTGPCLEF- 209
 QY 203 DNICIGTNTAG--KGAITAGTSNFSNNCDLFINNACAGAT-FSPICSLTGRGN 259
 Db 210 -----NNAAGGTAGGALFANACR-IENNSQPIYFLNOSGLGALRVHQECILTKVTGS 262
 QY 260 IVEYNNCFNVEFASSEADGAIKVTTRLDVTGNGRIFFSNITKKNGAIYAVVT 319
 Db 263 VIFENN--FMEDDISANHSSGAIYICIS-CSIKDNGCIAFNNTAROGAICITSLT 319
 QY 320 LVNNGPTFYFNINANNKGAIIYIDGTSNSKISADRAHAIIFENIIVTVNANGTSTANP 379
 Db 320 IQDSGPVYFTNN--QGTWGAIMLRQDCACTLFADQGDIIIFNNHFKDFTSN--HVSVC 376
 QY 380 PRNNAIVASSSGEITLLGAGSNNLITDPL-----EVSNAVSSTFKEDQDQGSVYFSG 435
 Db 377 TRNVSILTVGASG-----HSATFYDPLORYTIONS--IQKFPNDEPHLCTILFSS 425
 QY 436 -----ATVNSADF--HORNLOTKPAPLTLNSGFLCIEDHQLTVNFTGTGVVSLGNG 488
 Db 426 TYIDTSTSRDFTSHRN-----HGLYNGTLALEDRAEMKVVYKFDOGGGLRLIGSR 478
 QY 489 AVLSCKYKAGNSASNAITLKHIGLNLISLKSAGAEIPLLVEPTNNSNNYATDAATF 548
 Db 479 AVFSTTDEGSSSVGVININNLAINPLIL-GNRVAPKIMIRPTGSSAPYSEDNNEII 537
 QY 549 SLSDVKLSLIDYDGNSPRESTDLTHALSSOPMLSTSEASDQSLSDDDDFGLN-VPHYG 607
 Db 538 NLISG-PLSLDDENLIDPYDTADLAQPIAEVPLLYLLDTAHHINTDNEYPEGALNTQHYG 596

QY 608 WQGLMTWGMMAKTODPEPASSATITDPOKANRFHRTLLTLWLPAQIVSPKRRPLANTL 667
 Db 597 YGVVSPYVIEITITTSOTSS-----EDVTNLTLRQLYGDWPTGTGKVPENKKGIDIASAF 651
 QY 668 W---GNMLLATESLKNSNELTPSDHPFGITGGGIGGMVYODPRENHGPFMRSSGSAG 724
 Db 652 WQSFHNLFLATLRYOTQGOIAPT-----ASGEATRLVHONSNDKAGFMEATGTSLG 705
 QY 725 MIA--GQTHFSLKFSQTYTRKLNERYAKNNVSSKNVSCGEMLES-LOEGFLTLKLYGLY 781
 Db 706 TTSNTASNSHSCVNSOLFNSNIYESHSDNSVASHTTVALQINNWDLERSTASLA-Y 764
 QY 782 SYGDHCHHFFYTQGN--LTSQGTFRSOTMCGAVFYDPLMKPFGSTHILTLAPLCAICIT 839
 Db 765 SYSN--HHIKASGYSGKIQTEGKCYSTLLGALSCSLQ-WRSRPLHFPFIOALAVR 820
 QY 840 SLSHFTVGANPRSFSTKPTLINVLPIGVKCFMNAFORQATVELAYOPVLYROEP 899
 Db 821 SNOTAFQESGDKARKFSYHKPLYNLTPLUGIQSWESEKFLPTVYNIELATOPVLYXONP 880
 QY 900 GIATOLLASKGIWFGSSPSSSRHMSYKISQOTPLSMWTLHFOYHGFYSSSTPCNYLNG 959
 Db 881 EIVNSIESGSSWLLSGTLLANALATAFGRNOLFIFPKLSVFLDYOGVSSSTTHYLHA 940
 QY 960 EIALRF 965
 Db 941 GTTFKF 946

RESULT 8
 ID 09PL46 PRELIMINARY: PRT: 1025 AA.
 AC 09PL46: 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE POLYMORPHIC MEMBRANE PROTEIN E/F FAMILY.
 GN Chlamydia muridarum.
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGG:
 RX MEDLINE-20150255: PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Read T.D., Brunham R., Shen C., Gill S.R., Heidelberg J.F., White O.,
 RA Hickey E.K., Peterson J., Umayan L.A., Uterback T., Berry K.,
 RA Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
 RA Dodson R., Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G.,
 RA Salzberg S.L., Eisen J., Fraser C.M.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE002293: AAF39131.1: -;
 TR TIGR: TC0262;
 SQ SEQUENCE 1025 AA: 111458 MW: EA472E07818B9FE2 CRC64:
 Query Match 15.6%; Score 793; DB 2: Length 1025;
 Best Local Similarity 27.0%; Pred. No. 8.9e-46;
 Matches 269; Conservative 158; Mismatches 419; Indels 152; Gaps 35:
 QY 68 LQKTPNCGAVTTIDYLSFDTQKEGIFYFAKNLTPESGAIYGAASPNPTEIRDTIGPV 127

ID	992882	PRELIMINARY:	PRG:	934 AA.
AC	092882			
DT	01-MAY-1999	(TREMblrel. 10, Created)		
DT	01-MAY-1999	(TREMblrel. 10, Last sequence update)		
DT	01-OCT-2000	(TREMblrel. 15, Last annotation update)		
DE	POLYOMORPHIC OUTER MEMBRANE PROTEIN E FAMILY PMP_16			
GN	PMP_16			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae)			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-CML029.			
RX	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell J., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RT	Ollinger L., Gimwood J., Davis R.W., Stephens R.S., Trachomatis."			
RL	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
DR	Nat. Genet. 21:385-389(1999).			
SO	EMBL: AE001631; AADI8609.1; "			
	SEQUENCE 934 AA; 102149 MW; E88C8F6B1D3C47 CRC64;			
Query Match	20.7%;	Score 1052;	DB 2;	Length 934;
Best Local Similarity	31.2%;	Pred. No. 1,6e-63;		
Matches 296;	Conservative 158;	Mismatches 431;	Indels 64;	Gaps 27
35	PKRESLSNK-----ISLTGDFHNL-NCYLDLRYLAIILOKTPNEGAAVITIDYLSFF	87		
32	PLADSLSTLTGSPITTPYGNRHNSSODIVLSYKSIDNILLMTSAGCAVSCNFTL-L	89		
88	DIOKEIYFAKMLTPESGAIGYASPSPTVEIRDTTGPVIEFNNTCCRFPTSSNRNAV	147		
90	SNVEDHAFESKNIATIGTGALACOG---ACITTKNRGPIFFPSNRGLN-----NAST	138		
148	NKIREGAIHAO-NLYINHHDVVGPMKNFSYVRGASIFRANFVYSENOSCFELFMDNIC	206		
139	GGETRGCAIACNCDFTISONOGFFYVNVSVNMWGGALSTNGCHRIOSNAPLFFEN--	195		
207	IQENTAGKGAIFYAGTSNSFESNNCDLFFINNACAGALFSEI-CSLTGNGRIYFYNN	265		
196	---NTAPSGGALRSEVTTISDNTRPITYFNKCGNCGALQTSVYIAKKNSSGVIFENN	252		
266	RCFKNVETASSEASDGAIKVTRLDVYTGNGRRIFFSDNITKNGCAIYAPVYTLVNDGP	325		
253	TALSG--SINGSGGAI-YTNLSDIDNPCTILFENNCAIRGGAICTQFLIKRSH	309		
326	TYFINNANNKCAIYIDGTSNKSISMDRAIIFENYIVNVINANTSTISANPPRNAI	385		
310	VYFTNN-OGWVGALMLLDSTCLLEAGGQNIAPONNEVLTFTG-----RTNAI	358		
386	TVASSGEIILGAGSSONLIFYDPIEVSNAQVS-VSEFNKADOTGSVYFSGATVNSADFH	444		
359	HCFPNS-NIDLGANKGYTAPFDPIEQHPPTNPLFNPNANHQTILFESSAYIPEASDY	417		
445	QRNLQTPAPDLTNSGFLCTEDHAQLTVNRFOTGGVSVISGNAYVLSCKYNGAGNSAS-	503		
418	ENNFISSSKSTSELNCGVLSIEDRAGQFYKFTQGGILKGHAASLATTANSETPSTSV	477		
504	NASTLTHIGINSLILKSGAEIPLWVEPTNNNNYTAADIAAFSSLDVLSIDYGN	563		
478	GSOYIITNNLAINLPSIIAKG-KAPTLWIRPLQSSAPTEEDNPPITLISG-PLTILNEENR	535		
564	SPYESTDLTHALSSQPMLSISEASDNDLRSDMDFSGLN-VPHYGWGLMTGNAKTODP	622		
536	DVDSIDSLSEPLQNIHLTSLSDVTAARIINTDFNPRESENATEHYGOGIWSPYWET--I	593		
623	EPASSATITDQKANRPHRTLLTLWLPAGYVPSRKHNRPLANTLMGMLLATESLKNSA	682		
594	TTTNNAST--ETANTLVKALYANNTPRGIVNPNPYQODLTTPLQMSFHTMFSILRSYN	650		
683	ELTPSD--HPWMTGTGCGIMNVYODPRENNPCGFIMRSGYS--AGMTACQHTPTSLKFS	738		

D	b	651	RTGSDIERPFLLEIOGIADGLFVHONSIPGAHCPRIDISTGVSLQASSFTSLHOKISLIGFA	710
O	y	739	QTATKLNRYAKNNVSSKN--YSCOGEMLEFLDGEFLTATKLVLGYLSTGDNCHIFFTYOGE	796
D	b	711	QEFTRKEIGSSNNVAHNITVSSLVELELPW-FOENAF-ATSTVLVAGYGDDHLHLSPHSIQ	768
O	y	797	NITSQGTFRSOTMGCVAFFDLPMKPFCSHTILTAPELAGLGIYSLSLSHFTEVGAVPPRSFS	856
D	b	769	E-QAEGTCYSTHTLAALICSEFPWOOKSYLHL-SPEVOAIRSHQITAEEDIGNPKREV	825
O	y	857	TKTPLINLVPGVSGSEMNTAFORQATVELAYOPVLXHYRDEPGIATOLLSKSIWFSGS	916
D	b	826	SOKRPYYNI-TPLIGIOGGKKOSKFHVTEWTLELSTOPYLQONPOCYTLVLLASGGSWDILG	885
O	y	917	SPSSRHMSYKISOOTOPLSWLTLFHOXHGFYSSTFCNYLNGELALRF	965
D	b	886	INVYANALGYKHVNOTALFRSLDELVDYGVSSSTSTHHLOAGSTLKXF	934
R	E	RESULT	5	
I	D	09JSET	PRELIMINARY:	PRT: 934 AA.
A	C	09JSET		
D	T	01-OCT-2000	(TREMBLERel. 15, Created)	
D	T	01-OCT-2000	(TREMBLERel. 15, Last sequence update)	
D	T	01-OCT-2000	(TREMBLERel. 15, Last annotation update)	
D	E	POLYMORPHIC OUTER MEMBRANE PROTEIN E FAMILY.		*
G	N	PMF_16.		
O	S	Chlamydia pneumoniae (Chlamydophila pneumoniae).		
O	C	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
O	X	NCBI_TaxId=83558;		
R	N	[1]		
R	P	SEQUENCE FROM N.A.		
R	C	STRAIN=J138;		
R	X	MEDLINE=20330349; PubMed=10871362;		
R	A	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
R	A	Shida T., Ishii K., Hattori M., Kohara S., Nakazawa T.;		
R	T	*Comparison of whole genome sequences of Chlamydia pneumoniae J138		
R	L	from Japan and CwL029 from USA.*		
R	RL	Nucleic Acids Res. 28:2311-2314(2000).		
R	DR	EMBL: AF002546; DAA98673.1; -		
S	O	SEQUENCE 934 AA; 102163 MW; 9025BEC976933BDA CRC64;		
Q	U	Query Match	20.6% Score 1046; DB 2; Length 934;	
B	e	Best Local Similarity	31.1%; Pred. No. 4, 2e-63;	
M	a	Matches 295; Conservative 158; Mismatches 432; Indels 64; Gaps		27,
O	y	35	PTKESLSNK-----ISLQDTNHLT-NCYLDNLRYLATLQKTPREGAAVTTDVLSFF	87
D	b	32	PLDLSLTLTGFGSPITTFVGNRRNSSODIVLSNYKSIDNLTLMTSAGVASCNNFL-L	89
O	y	88	DTOKEGIFFKNLTPESGGAIGVAPSPFVELRDITGPVIFENNCCREFTSNNAAV	147
D	b	90	SNVEDHAFFSKNLAIAGACIAACOG----ACTTKNRGPLTFESSNGLN-----NAST	138
O	y	148	NKIREGAAIAAQ-NLYINHNDVYGFKNFSYVRGAISTANTFYVSSENSOCFLFDNDIC	206
D	b	139	CGEIRGAILACNDFTISONGTFFVFNNNSVNMKGALSTNGHCRIQSRAPLFFN---	195
O	y	207	IQTNTACKGAIYAGTSNSFESENCDL.FPINMACCAGCAIFSPL-CSITGRNIIFYNN	265
D	b	196	--NTASGGCALRESENTTISDNTRPIFYFNNGCNGNAGAIOTSJYAIKKNSSVIFENN	252
O	y	266	RCFINVETASSEASDGAIYVTTRPLDVTGKRGRIFEESDNTKNGGAIVPYATLVLDNGP	325
D	b	253	TALSG--SINGSGSGCAT-VYTMLSIDNBGTLLENNNVCIRDGCAICTQPLTIKNSGH	309
O	y	326	TYFINNIANKGCAIYIDGTSSNKSISADRHAIFENFIYTVTNANGTSTSANPPRNAT	385
D	b	310	VYFINN-QGMNGGALLMDLSDTCLLFLPDGGINAFQNNEVEFLTTGG-----RYNAI	358
O	y	386	TVAASGEITLGACSSONLIYPYDIENVSNAGV--VSFNKEADOTGSVPFSGAIYNSADF	444

```

Db      1 MKKLEFFVLIGSSILGTFREVPSPSLKLRPLNLPYHMTGLFFPKVNLIGDTHNLDYHLDN 60
Oy      61 LRYTLAILOKTPNEGAATITDYLSFPTQKGIYFAKNLTPESGAGIYASPNPTVEI 120
Db      61 LKCLLACLRPPEGAALFYVDYLGFSPTQKGIYFAKNLTPESGAGIYASPNPTVEI 120
Oy      121 RTIGPVIFENNTCCRPFTSSNPNAVNRREGGAIHAQNLINHHNHDVYGFKNESYVR 180
Db      121 HNTIGPVLENNTCRLMTOTDPEDEGKARREGGAIHADVYISNNQNLVGIKFNAYQ 180
Oy      181 GCAISTANTFVVENOSCFLEMDNICIOTNTAGKCAIYAGTSNFSFNCDLFIINAC 240
Db      181 GCAIS-ANTFAKEMKSSFLCINSCICQTKGCAIYVSTSGENNKKDLFIQNSG 239
Oy      241 CAGGAIFFSPICSLTGNRGIVFYNNRCFKNVETASSEADGAIYTTLDYTGNGRIF 300
Db      240 CAGGAIFFSPICSLTGNRGIVFYNNRCFKNVETASSEADGAIYTTLDYTGNGRIF 299
Oy      301 FSDNTKKNNGAIYAPVTVLVNCPPTYFINNIANNKGAIIYIDGTSNKSISADRAHIFN 360
Db      300 FSDNISRNGGAIHAPCLLVNGPFTYFNNTANITGCAIYITGTETSKISADHIAITFD 359
Oy      361 ENIVTNVNTANGSTSANPPRRNAITVASSGEILLGAGSSONLIFYDPIEVSNAGVSYS 420
Db      360 NNISANATNADGSSNTNPPHNAITMDNSAGIELGACKSONLIFYDPIQVTVNAGVTVD 419
Oy      421 ENKEADQGSVYFSGATVNASDFHQRNLOTKTPAPLITLSNGFLCIEDHQLVYNNRTQNG 480
Db      420 FKKDASQTCVYFSGATVNASDIQANLOTKTPAPLITLSHGLCIEDRQLVYNNRTQNG 479
Oy      481 GVVNSGNGVILCYKRGACN-----SASNASITLKHIGLNLSSILSKACIPLIPLW 532
Db      480 GIVAAAGNVISYOHSTIDATOTPTTTTDDASVTLNIGLNLPSILDDGAEMLIPLWE 539
Oy      533 PTN---NSNNYADTAATFSLSDVNLSDIDYGNSPVESTDLTHAASSQPMILISEASD 588
Db      540 PLSTTGNTTITSDTAASFSLNGATLSLIDEDGNSPYENTDLSRALYQPMILISEASD 599
Oy      589 NQLRSDMDPESGLNPPHYKMOGLMTGMAKKTODPREASATITDPOKANRPHRTLLTTL 648
Db      600 NQLOSSEMDPESGLNPPHYKMOGLMTGMAKKTODPREASATITDPOKANRPHRTLLTTL 659
Oy      649 PAGYVSPKHSRPLANTLTMGNMLTLESKNSAELTSPDHPHMGITGGGLGMVYQDR 708
Db      660 PAGYIYSPKHSRPLANTLTMGNMLTLESKNSAELTSPDHPHMGITGGGLGMVYQDR 719
Oy      709 ENHPGFHMSRSGYACMIAGOTHTFSLKFSQTYTKLNERAKNNVSSKNYSCQCEMLFSL 768
Db      720 KHPGFHMSRSGYACMIAGOTHTFSLKFSQTYTKLNERAKNNVSSKNYSCQCEMLFSL 779
Oy      769 QEGFLITKLVGLSYGDHCHHFTYTOGENLTSQGTFRSQTMGAVFEDLPKMPFGSTHLL 828
Db      780 QEGFLITKLVGLSYGDHCHHFTYTOGENLTSQGTFRSQTMGAVFEDLPKMPFGSTHLL 839
Oy      829 TAPFLGALCIVSLSHFTFEGAVPRSFSTKTPILINLVYIGVSGPMNATQRPQATVEI 888
Db      840 TAPFLGALCIVSLSHFTFEGAVPRSFSTKTPILINLVYIGVSGPMNATQRPQATVEI 899
Oy      889 AYOPVLVYROEPCIAIOLKASKIWEFGSGSPSRHANSYKISQOTOPLSWLTLHFOYHGY 948
Db      900 AYOPVLVYROEPCIAIOLKASKIWEFGSGSPSRHANSYKISQOTOPLSWLTLHFOYHGY 959
Oy      949 SSSITPCNYLNGEALRFE 965
Db      960 SSSITPCNYLNGEALRFE 976

```

```

RESULT 2
Oy2883 PRELIMINARY: PRT: 938 AA.
AC O92883:
DT 01-MAY-1999 (TREMblrel. 10, Created)

Query Match      22.0%; Score 1120.5; DB 2; Length 938;
Best Local Similarity 30.9%; Pred. No. 3.4e-68;
Matches 308; Conservative 163; Mismatches 418; Indels 109; Gaps 31.

Oy      6 FFILGNSLSGLAREVPSRIFLMPNSVDPPTKESLSNKISLTGDTNHLT-----NC 56
Db      12 FTFVLAN--EGQLPLETYITLSPEVOAPQ-----GFTHNQMODLAIYGNHND 59
Oy      57 YLDNLRYLAILOKTPNEGAATITDYLSFPTQKGIYFAKNLTPESGAGIYASPNPTVEI 116
Db      60 FLDYKYY-----RSNGALITCKNLL--ISENIGNVFEXNVCPSNGAL-YAQNOC- 108
Oy      117 TVEIRDTIGPVIFENNTCCRPFTSSNPNAVNRREGGAIHAQNLINHHNHDVYGFKNF 176
Db      109 -----TIK--NQNVAFTNLVSDNPTTAGSL--GGLAFINCSITNNDGQGFVDL 159
Oy      177 STVRGCAISTANTFVVENOSCFLEMDNICIOTNTAGKCAIYAGTSNFSFNCDLFI 236
Db      160 ALNKGCAIYTFNLISIKDKKPIIHKONRAL--NSDSLGGIYSGNSLNIIEGNSAIOIT 217
Oy      237 NNACGAGAIFFS-PICSLTGNRGIVFYNNRCFKNVETASSEADGAIYTTLDYTGNGRIF 295
Db      218 SNSSGSGGIFSTQTLTJSSNKKLIEJSENSAFAN--NYGSNFPQGGCLTTTFTILNN 275
Oy      296 KGRIFSDNITKNYCGAIYAPVTVLVNCPPTYFINNIANNKGAIIYIDGTSNKSISADRAHIFN 352
Db      276 REGVLEFNNGOSNGCAIHAKSIIIEKNGPYFLNNTARGALLNLWSGNSGNSPILSA 335
Oy      353 DRHAITFENIVTNVNTANGSTSANPPRRNAITVASSGEILLGAGSSONLIFYDPIEV 412

```


N100_YEAST
ID N100_YEAST STANDARD: PRT: 959 AA.
AC Q02629;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).
GN NUP100 OR NSP100 OR YKL068W OR YKLJ36.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93054906; PubMed=1385442;
RA Wente S.R., Rout M.P., Blobel G.;
RT "A new family of yeast nuclear pore complex proteins.";
RL J. Cell Biol. 119:705-723(1992).
RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN=5288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins.";
RL Yeast 10:569-574(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC -1- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC -1- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z15035; CAAT8753.1;
CC DR EMBL: X75780; CAAS3406.1;
CC DR EMBL: Z28068; CAAB1905.1;
CC DR PIR: B44402; B44402.
CC DR PIR: S39173; S39173.
CC DR PIR: S44518; S44518.
CC DR SGD: S0001551; NUP100.
CC KW NUCLEAR protein; Transport; Repeat.
CC FT DOMAIN 33 571 29 X 6 AA APPROXIMATE REPEATS OF
CC FT G-L-F-G.
CC SQ SEQUENCE 959 AA; 99988 MW; D3985F9901BBA51 CRC64;

Query Match 2.8%; Score 141.5; DB 1; Length 959;
Best Local Similarity 19.8%; Pred. No. 0.3;
Matches 151; Conservative 81; Mismatches 257; Indels 273; Gaps 38;

OY 96 FAKULTESPSCAIG-----YASPNSTVEIRDTIGV-----IFENNT 133
DB 68 FGNNTT-ONNAFGQMGATONSPFGSLNSSNASNGTFFGSSSSMGSGFGNTNNAFNNS 126
OY 134 CCRPFTSSNPAANVKIEGGAIAHONLYINIHNDVGFPMKFNFSYVRGAISTANTFEVS 193
DB 127 -----NSTNSPFGNKKPRTG-----TLFGSONNNSAGTSSLSF-----GGGSTITTGFG 171
OY 194 ENOSCFLEFMDNICIOTNTAGKGAIVACTSNSPESNNCDLFFINNACCA----- 242
DB 172 NTGSSF-----GTGLNGGNSNIF-GAGNNSGNTTGLSFLGNOSSAFGTINNQGSLF 222

OY 243 -----GGAIIF--SPI--CSLTGNR-----GNIVFNYNRCFKVETASSE 277
DB 223 GQGSQNTNNAFGNQLGSSFGSKFVGSGSLFGQSNNTLGTNTNRNGLFGQMNSSNGC 282
OY 278 ASDGAI-----KVTTRLDVYGNRGRIFF--SDNTITNKYCAIYAPVTLV 321
DB 283 SSNSGLFGQNSMNSTQGVFGQNNQMOLNGNNNNSLFGKANTFSSASGGLFGQ--NNQ 340
OY 322 DNGPTTFINNANKGCAIYIDGTSNK-----I 350
DB 341 QOCGSLFGQNT--SQTGSSSLFGQNNQKQPNFTQSTNTGLFGQNNQOQOSTGLFGAK 398
OY 351 SADRHAIIFENENTVNTVANGSTANPPRRVAIVASSGEILLGASSONLIF--YD 408
DB 399 PACTTGSGLFGQNSTQPNLSFGTT--NWFTSN--TOSQGGSLGATKLTNMPGCGNP 452
OY 409 PIEVSNAGVSVSPNKADQTSVYFSGATVNSADFHQRNLQTKTPAPLTLNGLICIEDH 468
DB 453 TANQSGNSLFGTKPASTGSL-----FGNNTASTVPS--TNGL--FGNN 495
OY 469 AOLTVNRFTQTG-----GVYSLGNGAVLSCKNGAGNSASNASI-----TL 509
DB 496 ANNSTST-TNTGLFGAKPDSQSKPALGGGLF-----GNSNSNSTTIGQNKPVGCGTT 546
OY 510 KHIGL-----NLSTILKSG-----AEIPLWVEPTNN--SNVYTDATATFSLSDV 553
DB 547 QNTGLGATGTSSAVGSGTGLFGQNNNTLVNCTONVPPNNTTQNALGTTAVPSLQQA 606
OY 554 -----KLSTIDYDGNSPYEST-----DLTHA--LSQPMLSISEASDNO 590
DB 607 PTNQLFSLKISLPSNSTN-PVKATTSKVNAADKRNLSLSAVRLAKPLFASSNGDAK 665
OY 591 LRSDMDFSLNVPHTGMOGLMTWGAKT--ODPEASSATTDDOKANRFRITLLTW 647
DB 666 FOK-----MCKTLERSRGSSTNSITDESSYLSNDDL-- 700
OY 648 LPAQVPSPKHRSPLNLTLMGNMLLATESLKSNAELTPSDH 689
DB 701 -----FDPDKRRLKHLV-----IKNNKLNVLNH 724

RESULT 25
YN96_YEAST STANDARD: PRT: 1117 AA.
ID YN96_YEAST
AC P33753;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHEICAL 121.1 KDA PROTEIN IN BIO3-HXT17, INTERGENIC REGION
DE PRECURSOR.
GN YNR067C OR N3547.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX Duysterheft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YLR144C AND S.POMBE SPAC2303.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z71682; CAAG6349.1;
CC DR SGD: S0005350; YNR067C.
CC KW Hypothetical protein; Glycoprotein; signal.


```
QY 593 ---SDMPFSGINPHYMGGLMTN-----GNAKTOD----- 621
Db 1185 KTGCGDASFTLGNTG--GFVDLGTIEYVLKSDGNSNMWLTNDVKNPDPPIPNKRPDPKRD 1242
QY 622 -----PEPASSATITDPKANKFHRTLLLTWLPAGVYNSPKHR-SPLIANTLWGNL 672
Db 1243 KPDPNPKPDPPTPT-----PYPERITPSTAVL--NM- 1277
QY 673 LATESLXNSAEI-----TPSDHPFWGIT-----GGGL-----GMVY 704
Db 1278 AATLPVLPDAELNSIRERLINIKKASPHNNNVWGATYNTFRNNVTTDAGAGFEQTLGNTVG 1337
QY 705 QDPRENHGFHNRSSGYS-CHIAQOTHTFLKESQYTKLMEPRAKNNVSSKNISCGE 763
Db 1338 IDSRNDIP-----EGITTLGAFMGYSHS-HIGF-----DRGHSVGS--YSLGXY 1380
QY 764 MUFSLQEGFLTKLVGLSYGDHNGHFEYTOGENTLSQTFPSQTMGCAVFPDLPPKPPG 823
Db 1381 ASMEHESGFYLDGVYKLNRFKSNVAGKSSSG--AANGSYNSGLGHI--ETGKRFTD 1435
QY 824 STHILTPALGALGIYSSLSHPT 846
Db 1436 GNMNLTLP-----YASLTGFT 1450

RESULT 17
CSG_METFE STANDARD: PRT: 593 AA.
ID CSG_METFE
AC P27374;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN PRECURSOR (S-LAYER PROTEIN).
CN SIGA.
OS Methanothermobacter feravidus.
OC Archaea: Euryarchaeota: Methanobacteriales: Methanothermaceae:
OC Methanothermobacter.
OX NCBI_TaxID=2180;
RN 11
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-42.
RC STRAIN-DSM 2088 / V245;
RX MEDLINE-91293115; PubMed-1712296;
RA Broeckl G., Behr M., Fabry S., Hensel R., Kaudewitz H., Biendl E.,
  Koenig H.;
RT "Analysis and nucleotide sequence of the genes encoding the surface-
  layer glycoproteins of the hyperthermophilic methanogens
  Methanothermobacter feravidus and Methanothermobacter sociabilis.";
RL Eur. J. Biochem. 199;147-152(1991).
RN 12
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-94086483; PubMed-8262914;
RA Kather U., Schroder H., Haslinger E., Altmayer G., Schreiner R.,
  Wietand F., Haselbeck A., Konig H.;
RT "Primary structure of the heterosaccharide of the surface glycoprotein
  of Methanothermobacter feravidus.";
RL J. Biol. Chem. 268;26821-26826(1993).
CC -I- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
  OF PROTEINS WHICH COAT THE SURFACE OF THE CELL.
CC -I- SUBCELLULAR LOCATION: CELL WALL. THIS ARCHAEA IS COVERED BY A
  S-LAYER WITH HEXAGONAL SYMMETRY.
CC -I- PTM: PROTEIN CONTAINS ABOUT 10 N-LINKED GLYCANS WHICH CONSIST OF
  METHYL-MAN, MAN AND GALNAc RESIDUES IN A MOLAR RATIO OF 2:3:1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58297; CAA41230.1;
DR PIR: S16225; S16225.
```

```
KW Glycoprotein; Cell wall; S-layer; Signal.
FT SIGNAL 1 22
FT CHAIN 23 593
FT CARBOHYD 29 29 N-LINKED (GALNAc . . .)
FT CARBOHYD 58 58 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 66 66 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 74 74 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 114 114 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 122 122 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 145 145 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 148 148 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 158 158 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 176 176 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 208 208 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 231 231 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 326 326 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 336 336 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 340 340 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 431 431 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 471 471 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 500 500 N-LINKED (GALNAc . . .) (POTENTIAL)
FT CARBOHYD 516 516 N-LINKED (GALNAc . . .) (POTENTIAL)
SQ SEQUENCE 593 AA; 65481 MW; E06466967B7611C CRC64;

Query Match 3.0%; Score 153.5; DB 1; Length 593;
Best Local Similarity 22.9%; Pred. No. 0.029;
Matches 119; Conservative 46; Mismatches 189; Indels 165; Gaps 28;

QY 92 EGIYFAKNLTPESSGA--IGVASPNSPYEIRDTIGPVIFENNTCCRPSTSSNPAAYNK 149
Db 61 EMLVKNLSLTKNSGSAITINAVSSEKSTITK-----ANNWIEGF----- 101
QY 150 IREGG--AIHAQNL--YINHNDDVGFPMKNFSYVYGAI-----STANTF--VVSSENS 197
Db 102 IIGCKNGIYEMENVGCTITNTTIOAFVSGMEYVGGNIGCLVNSNNITNNIIRNTW 161
QY 198 CFLFMDNIC-----IQTYTAGKG--AIYAGTSNFSFESN-----CDLF- 234
Db 162 NC1---NVCEKGNIIKNTIMVSGIGIYVWGFNFKENNIENNRIINATYGCITFLR 218
QY 235 -----FINNACCAGAIIFSPIC-----SLTGNRGNTV 261
Db 219 PSNNKTCRNVIYAVSSGGGSGGSGAICIDVSDINIVKDNIGINCDDGLFTDGMIGNE---- 274
QY 262 FYNNRCFKNVETASPSASDGAIKVTRIDVYGNRGRIFFSDNITKNYGAATAPVYTLV 321
Db 275 -ITNIFKNCQVAVSESTYGPA-----SRNNKIY--GNFYINVEFTALSDPGELV 321
QY 322 DN-----GPTFINNIANNKGAIIYDGTSSKISADRAHAIIFN-----ENIYT 365
Db 322 DNIMNTTEGGNWSNNTGNNTG----DGTGNIPYYDKPLVDAIEDIAKPSGLEV 376
QY 366 NYTNANGTSTANPPRRNITVAASSGEIL-----LGAGSSONLIFYPIEVSAGVSV 419
Db 377 RKKNLGRADIKIDPLTKIKIRISCDNDVYEYEIFDLSAGEO-IYRWKI-VPEENHII 434
QY 420 SPNKADQTSVYFVSGATVNSADFHQRNLQTKTPADLTLSNGLIEDHQAQLVNRFTQT 479
Db 435 KAEIPYSAGYIL--GTNIRADI--SNVEFSK-----IVQGF-----YQKFTT 475
QY 480 GGVSLSGNQAV-----LSCYK--NGAGSASVASTLK 510
Db 476 IYLTNIGKSTIITIKYIISITVPVNTKVSYRELITLK 514

RESULT 18
CSG_METSC STANDARD: PRT: 593 AA.
ID CSG_METSC
AC P27374;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
```



```
Db 1614 -DNVEGSNLTDAADITSTSVVTAAGSTDAAGNDVATMS-----KN 1654
QY 620 QDPEPASSATITDPOKANRHRITLLTWLPAGYVSPKHSRPLANTLMGNMLATES-- 677
Db 1655 AYDVATDASVNDIAKA-----LDAGYTNNEFLTSLNGTTELNSALKOVSGS 1703
QY 678 -----LKMSAELTPSDHPWGLTGGGLGMVYQ--DPR-----NHGQFHM----- 716
Db 1704 QATTVREARVLNRFNSMLAADAP--KVNGLAFFNVAKDPRRELGNTEYDLALIRK 1760
QY 717 -----RSSGYSA-----GMIAGOTHTFSKFSQYTK-LN-----ERY 748
Db 1761 TIDSESGTMSLEVCIALDDGAKAKDNCVGTGYSQFGLKHMSPDNCKMNNMNLARY 1820
QY 749 AKNNYSK-----NYSQGEMLFSLOEGFLTTLKVLGYLSYGDHN 787
Db 1821 DVHNLDSRSIAFGCNTKNTADTVKQOYLEFRSEGAKEFEESEGLKVPYAGV-----K 1874
QY 788 CHHFTYQGENLTSGTFSQTMGCAVEFDLPKMPGSGTHI-----LTPAPLIGALGYSS- 841
Db 1875 LRH-----TLEGGYQERNAGD--FNLNNNSGSETAVDISYGLKLDYAGKDGWSASA 1923
QY 842 -----LSHFE-----VGAYPEFST-----KTPLIINVLPICVGS 873
Db 1924 TLEGGPNLSYAKSORTASLAGAGSOFHFNVDGQKCGINSLTSGVKYS 1972

RESULT 10
190K RICRI STANDARD: PRT: 2249 AA.
ID 190K RICRI
AC P15921:
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-OCT-1994 (Rel. 30, Last annotation update)
DE 190 KDA ANTIGEN PRECURSOR (CELL SURFACE ANTIGEN).
OS Rickettsia rickettsii.
OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales:
OC Rickettsiaceae: Rickettsiidae: Rickettsia.
OX NCBI_TaxID=783:
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=R:
RX MEDLINE=90354033: PubMed=2117568:
RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.:
RA "A protective protein antigen of Rickettsia rickettsii has tandemly
RT repeated, near-identical sequences."
RL Infect. Immun. 58:2760-2769(1990).
CC - FUNCTION: ELICITS PROTECTIVE IMMUNITY.
CC - PTM: GLYCOSYLATED (POSSIBLE).
CC - DISEASE: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PATHOGENESIS OF
CC AND IMMUNITY TO ROCKY MOUNTAIN SPOTTED FEVER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31227: AAA26380.1:
DR PIR: A41477: A41477
KM Antigen: Repeat: Signal: S-layer: Glycoprotein.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2249 190 KDA ANTIGEN.
FT DOMAIN 212 1180 13 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 358 A (TYPE I).
FT REPEAT 359 430 B (TYPE II).
FT REPEAT 431 505 C (TYPE II).
FT REPEAT 506 577 D (TYPE I).
FT REPEAT 578 652 E (TYPE II).
FT REPEAT 653 724 F (TYPE I).
FT REPEAT G (TYPE II).
```

```
FT REPEAT 725 799 H (TYPE I).
FT REPEAT 800 874 I (TYPE I).
FT REPEAT 875 949 J (TYPE I).
FT REPEAT 950 1021 K (TYPE II).
FT REPEAT 1022 1093 L (TYPE II).
FT REPEAT 1094 1165 M (TYPE II).
FT REPEAT 1166 1180 TYPE I (INCOMPLETE).
SO SEQUENCE 2249 AA: 224333 MW: A9D6646C089DF087 CRC64:

Query Match 3.4%: Score 174: DB 1: Length 2249:
Best Local Similarity 21.2%: Pred. No. 0.0098:
Matches 164: Conservative 90: Mismatches 294: Indels 226: Gaps 34:

QY 10 IGNSLSGLAREPSPHIFLMPNSVDPPTKESLSKIS-----LTGDTNHLNLTCLD 59
Db 212 IGNTNALATVWVGATLGGAVIKATTTKLTNASVLTITNANAVLTGALDNTGG--D 269
QY 60 NLRYL-----ATLQTPNEGAAVTITDYLSPDQKEGIVFAKNLT---PESGALGVASP 113
Db 270 NGVVLNGLALSQVTDGIGNTNSLATISVGAGTATLGGAVIKATTTKLTDASAVKFTNP 329
QY 114 NSPTVEIRD-----IGPVIFE-NNTCCRPFTSSNPNAAVNRIR-----EGCAIHAQNYI 163
Db 330 VVVTGALDNTGNANNGIYVTFNGNSTVTCNVGNTNALATVWVGAGILQYGGVAVANTINL 389
QY 164 NIINIDVGFPMKNFSYVRGGAISTANTFVSENOGCFLEMDN-----ICIQ 208
Db 390 TDMASAVTF-TNPVVVTAIDNTGNA-----NNGIVFTGNSVYTDIGNTNALATVWVG 443
QY 209 TNTACKGATVAGTNSPESNNCDLFFINNACGAGAFS-----PICSLT 254
Db 444 AGTATLGAIVIKATTTKLTNASVLTITNANAVLTGALDNTGGDNVGLNGLALSQVT 503
QY 255 GNRGNIVEYNNRCFNKVEATSEASDGA-IKVTT-----RLDVT 293
Db 504 CNIG-----NTNSLATISVGAGTATLGGAVIKATTTKLTDASAVKFTNPVVVTAIDNT 558
QY 294 G--NGRIFFSDNT-----TKNNGALYAPVTLVNDGPIYFINN 331
Db 559 GNANNGIYFTGNSVYTDGIGNTNSLATISVGAGTATLGGAVIKATTTKLTNASVLTIT 618
QY 332 JANNNGAIYIDGTSNSKISADRHAIIFENIVTWTNANGSTSANPRRATVASS 391
Db 619 NNN-----AVLTGALDNTGGDNVGLNGLALSQVTGIDGNTNS-----LATISGA 666
QY 392 GEILLAGSSONLIFYDEIENSNAVSSEFKEA-----DOTGS-----VVFSGATVNS 440
Db 667 GTATLGA-----VIKATTTKLTNASAVKFTNPVVVTAIDNTGANNNGIYFTGNSVY 722
QY 441 ADFHQRNL-----QTPPAPLISNGFLCIEDHQAULTVRF 476
Db 723 GDIGTNTLATVWVGATLGGAVIKATTTKLTNASVLTITNA-----NAVLTGAID 776
QY 477 TQTG-----GVSLGNGAVLSCKKNGASNASASTLKHIGLINSILKSGAEPLLWVE 532
Db 777 NTTGGDNVGLNL-NGA-LSQVTDGIGNTNSLATIS--VAGATVTL--GGAIVIKATTTK 829
QY 533 PTN-----NSN-----NYT-ADTAATFSLDVYKSLDIDYGS-----564
Db 830 LTNASVLTITNANAVLTGAVDNTGGDNVGLNGLALSQVTGIDGNTNSLATISVAG 889
QY 565 -----PIESTDLTALSSQPLMSISEAS-----DNQLRDDDDFQSLN 602
Db 890 TATLGAIVIKATTTKLTNAS--VLTITNANAVLTGALDNTGGDNVGLNLN 940

RESULT 11
VAC3_HELPY STANDARD: PRT: 1310 AA.
ID VAC3_HELPY
AC Q48253:
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

```

Db 1050 -LSMSTSMIGTTTANNNGVTYVIGMARVNIIGSDIPVASFRTGSDSAGLQGNYSQVI 1108
OY 442 DFHQRNIQTTPAPRLTSLNSGFLCEDHQAOLVNNRFTQTGGVNSLNG-----AVLSCKYNG 497
Db 1109 DFGTYNNLGI--VNSNIILGGTTAINKIDLVNTLTFLFASGISTWNCNNTSIETTLTLANGN 1167
OY 498 AG-----SSASMASTTEK-----HGLNLSILKSGAEI-----PLWPEP 533
Db 1168 IGHVILEGAQVAVTTTGTGTTILVQDNANANFSGTYOTYTLILGGARFNGTSLSPFAVYG 1227
OY 534 TNSNNYTTADTAAT-----FSLSDVLSLIDYGNSPY-----ESTDLTALSSQPM 580
Db 1228 SNRVRVNTSLIRANODVYITRTNNAENVVYTNDIANSFPGCARGVQDQNTTFVNAITNAAAY 1287
OY 581 LSISSADNDQLRSSDMDPSGLNVPNHYGMOGLTWGMAKTQDEPPASSATITDPQ----- 634
Db 1288 NNLLAKNS--ANSAFVGAIVTD-----TSAITVWQDLDAKD 1324
OY 635 -KARFRRTLLTLPAGVYPRPKHNSPLANTLNGMMLLATESIKNSAEILTPSDHPWG 693
Db 1325 IOAOLGNRLGALRYLG--TPETAEMAGPREGATISAAVAAGEALIDNVA-----YG 1372
OY 694 ITGGGLGMVYQDREHNPFGHRRSGYSAGM--IAGQTHFESLKFSSQTYTKL-NERYAK 750
Db 1373 IMAKPFYTDAAHQSKKGLAGTKKATGTYVIGDTLTANDMLMIGAALGKTIKIRKQDKK 1433
OY 751 NNVS-----KNYSCGEMLESLQESFLLTKLVGLSYGDHNCHEFTYG 795
Db 1433 GDKTDVNGFSFLYGAQOVLKNFFAQSALFESLNO-----YANKSQRFFEDAN 1480
OY 796 ENLTSQ---GFFRSQYTG--AAVFDLPKMPFSTHILTAPFLGALGILYSSLHTEGA 850
Db 1481 GNMSKQIAGAHYDMTGTGMLTYGYDYN---AMQGLVTPMAGLSYLSKSDENKKEGT 1536
OY 851 -----YPRSESTPLINLVPIGVGSEPMNATQRPQAMVELAYQPLVYR----- 896
Db 1537 TVANKQVNSKTSDDTLI---YQAKVAGSTMNT-----DLAYPEVNAFVHKHYKG 1585
OY 897 -----OEPGIATOLLASKGIWFGSGSPSSRHMSYKTSQQTQLSWLT 939
Db 1586 RLSTQSVLDGQVPTCINOPDRTTKTSYNGL---SASIRSKMEYGIQYDAQISSKYT 1644
OY 940 LH 941
Db 1643 AH 1644

RESULT 8
OMP_B_RICJA STANDARD: PRT: 1656 AA.
AC 006653:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)
DE (SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCAS) (ROMP B)
DE (CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN) (SURFACE PROTEIN ANTIGEN)
DE (120 KDA OUTER MEMBRANE PROTEIN OMPB): 32 KDA BETA PEPTIDE).
GN OMPB.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiaceae; Rickettsia.
OX NCBI_TaxId=35790:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YH:
RA Uchiyama T.;
RT Sequencing of the gene encoding the protein romp B of Rickettsia
RT japonica.
CC Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BT

```

```
CC SIMILARITY)).  
CC -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR  
CC (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
CC LAYER WITH HEXAGONAL SYMMETRY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AB003681; BAA20138.1;  
KW Antigen; S-layer; Cell wall. 120 KDA SURFACE-EXPOSED PROTEIN.  
FT CHAIN 1 1338  
FT CHAIN 1339 1656 32 KDA BETA PEPTIDE.  
FT DOMAIN 528 533 POLY-GLY.  
FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
SQ  
  
Query Match 3.7%; Score 189; DB 1; Length 1656;  
Best Local Similarity 19.2%; Pred. No. 0.00076;  
Matches 218; Conservative 129; Mismatches 428; Indels 358; Gaps 45;
```


OY		156	IHHONLYINHNHVVCEFMKPNFSVAGCAGAISTANFPEVEENOSCFLEMDNICIOJNTA--G	213
Db		115	IFKENSTLMLDYL-I-FSGNVAAGYGGAITSSET---NDTGAVDLRYTNAAFRNNINDG	170
OY		214	KGGAITAGTSNSFNESNCKDLFEINNACCAGAIFFSPICSLTGNRCGNIVFYNNRCKRVET	273
Db		171	KGGAIV-----LTNDDYVL-----SDVIPDNQAY---TS	197
OY		274	ASSEASDGCAIKTTTLDTGR-GRIFEGDNIKNKGALY----AP	316
Db		198	TYSVDGDGGAI-----DYTDNKSOKHPISGYTIYNNTAFNTAEYGGAITYNSTAP	251
OY		317	VWTVLVNDGPYYFINN---IANKNGCALIIDCTSN-----SKISADHAATIFMENI	363
Db		252	YLLDISVDISOYNGGVLDENNNSAAGYCDGPSNAAGGFMYLGSEVFED---IADKTL	308
OY		364	VTVNVANGTSTISANPPRRNATIVASSGEILLGGSSONLIIFYPIEVSNAGSYSFENK	423
Db		309	VIGNTENDGADVDSI---AGTGLTKFKGSODLVLANMND--FTPGEMDIENGCEVALRSN	362
OY		424	EADQTGSVPS-----GATVSADEFHORNL-----OTKPAPPLLSNGFCIEDH	468
Db		363	SIMNWGTDCQQDPQCGLTGITGISIDYOYNQMELNWGSTOOTFIHALTFONGRLINDAG	422
OY		469	AOLTVNRFTOTGCVSVSLCGNAVLCSCYKNCAGNASASNASTLKHKILGNLSILKSGAELP.	528
Db		423	GNTVNQGSPAGIIIEGAGQ---LTIANQNSYVLAGNQSMALTG-----DIVDGAVALS.	474
OY		529	LWNEPTNNSNNTYADTPRAATFSLSDVKLSLIDXYG-----NSPFESTULIALSQPMLS	582
Db		475	-----BGDAADLTALADDPOSITVLNGVGVLIDSDFSTWS-----GTSYNDGLE	517
OY		583	ISEASDNQLRSD-MDFSGLNPVPHYMGOG	610
Db		518	VSGSGTGIVGSDPVLDLAGGDNHIIGDQ	546
RESULT		4		
OMP_B_RICPR		ID	OMP_B_RICPR	STANDARD:
AC		053020:		PRT: 1643 AA.
DT		30-MAY-2000 (Rel. 39,		Last sequence update)
DT		30-MAY-2000 (Rel. 39,		Last annotation update)
DE		OUTER MEMBRANE PROTEIN B PRECURSOR (168 KDA SURFACE-LAYER PROTEIN)		
DE		(SURFACE PROTEIN ANTIGEN) (CELL SURFACE ANTIGEN) (SCA5) (ROMP B)		
DE		[CONTAINS: 120 KDA SURFACE-EXPOSED PROTEIN (SURFACE PROTEIN ANTIGEN)		
DE		(120 KDA OUTER MEMBRANE PROTEIN OMPB); 32 KDA BETA PEPTIDE].		
CN		OMP_B OR SPA OR SPA OR RP704.		
OS		Rickettsia prowazekii.		
OC		Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;		
OC		Rickettsiaceae; Rickettsiae; Rickettsia.		
OX		NCBI_Taxid=782;		
RN		{1}		
RP		SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC		STRAIN-BREINL:		
RC		MEDLINE=91045972; PubMed=2122457;		
RA		Carl M., Dobson M.E., Ching W.M., Dasch G.A.:		
RT		"Characterization of the gene encoding the protective paracrystalline-		
RT		surface-layer protein of Rickettsia prowazekii: presence of a		
RT		truncated identical homolog in Rickettsia typhi."		
RU		Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).		
RN		{2}		
RP		SEQUENCE FROM N.A.		
RC		STRAIN-BREINL:		
RA		Morón C.G., Yu X.J., Walker D.H.:		
RT		"Sequence analysis of ompB of Rickettsia prowazekii.";		
RL		Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RN		{3}		
RP		SEQUENCE FROM N.A.		
RC		STRAIN-MADRID F.:		
RC		MEDLINE=9903499; Pubmed=9823893;		
RX		Andersson S.G.E., Zomorodipour A., Andersson J.O.,		

```

RA Eriksson A.-S., Winkler H.H., Kurland C.G.:
RT "The genome sequence of Rickettsia prowazekii and the origin of
RL mitochondria."
RL Nature 396:133-140(1998).
RL
RL PARTIAL SEQUENCE.
RP [1]
RP
RC STRAIN=BREIN;
RC
RX MEDLINE=92114896; Pubmed=1370573;
RA Ching W.M., Carl M., Dasch G.A.:
RT "Mapping of monoclonal antibody binding sites on CNR fragments of
RT the S-layer protein antigens of Rickettsia typhi and Rickettsia
RT prowazekii."
RL Mol. Immunol. 29:95-105(1992).
RL [5]
RP IDENTIFICATION OF CLEAVAGE SITE.
RP
RX MEDLINE=92104668; Pubmed=1729180.
RA Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.:
RT "Evidence for proteolytic cleavage of the 120-kilodalton outer
RT membrane protein of rickettsiae: identification of an avirulent mutant
RT deficient in processing."
RL Infect. Immun. 60:159-165(1992).
CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC LAYER WITH HEXAGONAL SYMMETRY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37647; AAA26390.1; ALT.INIT.
DR EMBL: AF161079; AAD4234.1; -
DR EMBL: AJ235273; CA15140.1; -
KW Antigen; S-layer; Cell wall.
FT CHAIN 1 1328 120 KDA SURFACE-EXPOSED PROTEIN.
FT CHAIN 1329 1643 32 KDA BETA PEPTIDE.
FT VARIANT 257 257 V -> A (IN STRAIN BREINL).
FT VARIANT 1010 1010 Y -> D (IN STRAIN BREINL).
FT VARIANT 1450 1450 A -> S (IN STRAIN BREINL).
FT VARIANT 178 179 AA -> VC (IN REF. 1).
FT CONFLICT 191 201 TTQEAFLPGA -> INSRSSSYHLVS (IN REF. 1).
FT CONFLICT 212 212 T -> I (IN REF. 1).
FT CONFLICT 313 313 Q -> L (IN REF. 1).
FT CONFLICT 1104 1104 D -> G (IN REF. 2).
FT CONFLICT 1123 1123 T -> S (IN REF. 2).
SQ SEQUENCE 1643 AA; 169854 MW; 735FDf392E634bCC CRC64;

Query Match 3.8%; Score 193; DB 1; Length 1643;
Best Local Similarity 19.1%; Pred. No. 0.00043;
Matches 192; Conservative 133; Mismatches 434; Indels 244; Gaps 37

QY 41 SNKISLTGDT-----HNLTNCYLDNLRYLAIALOKTPNEGAAYVITDYLSFFDTQKEG 93
DB 633 SKTILNAGVAINELVLENNQSVQQLNHNHTYLLITKTIANAQOQIIVAADPLNTNTLLAG 692
QY 94 IYFAANLPBESGALGYSAPNSPYEIRITIGRYVFENNTPCCARPTFSSNPNAAVKNIRRG 153
DB 693 TNLGSAENPLSTIRPATYKAANADSI-LVNGKGVNLIANN-----ITTDANNGSLAFRSG 746
QY 154 GAI-----HAONLYINHNDVGVGFMKNFSYVRGAISTANTFVVSENQSC---- 198
DB 747 GSIYSGTVGGQGGHKLNNLLIDNGTGVVFLGDTTENGKTIKGRKSIILDISNYYTDHVE 806
QY 199 -----FLEMDNICIOTNTAGKCGALVAGTSNFSFESNCDLFF--INNACCGAIAIF 247

```

QY 181 GGAISTANTFVSENOGCFLEMDNICIOTNTAGKGAIIYAGTSNFSFNCDLFFINAC 240
 DB 180 GCAISTANTFVSENOGCFLEMDNICIOTNTAGKGAIIYAGTSNFSFNCDLFFINAC 239
 QY 241 CAGAGIFSPICSLGNRCNIVFYNNRCKNVEYASSAASDGAIKVTRRLDVTGRGRIF 300
 DB 240 CAGAGIFSPICSLGNRCNIVFYNNRCKNVEYASSAASDGAIKVTRRLDVTGRGRIF 299
 QY 301 FSDNITKNGGAIYAPVTVLVNCPFTYFNNIANNKGAIIYIDGTSNKSISADRAIIIFN 360
 DB 300 FSDNITKNGGAIYAPVTVLVNCPFTYFNNIANNKGAIIYIDGTSNKSISADRAIIIFN 359
 QY 361 ENIVTNTNANGTSTSNAPPRNATIVASSGCEILLGAGSSONLIFYDPIEVSNAVSYS 420
 DB 360 ENIVTNTNANGTSTSNAPPRNATIVASSGCEILLGAGSSONLIFYDPIEVSNAVSYS 419
 QY 421 FNKEADOTGVSFSGATVNSADFHORNLOTPAPLTLISNGCLIEDHAOLTVNFTOTG 480
 DB 420 FNKEADOTGVSFSGATVNSADFHORNLOTPAPLTLISNGCLIEDHAOLTVNFTOTG 479
 QY 481 GVSISLGNCAVLSCKYKNGAGNSASNAITLKHIGLNLSSILKSGAEIPLLWEPITNSNNY 540
 DB 480 GVSISLGNCAVLSCKYKNGAGNSASNAITLKHIGLNLSSILKSGAEIPLLWEPITNSNNY 539
 QY 541 TADTATATSLSDVKLSLIDDYGNSPYESTDLTHALSSQPMLSISASDQNLASDMDTSG 600
 DB 540 TADTATATSLSDVKLSLIDDYGNSPYESTDLTHALSSQPMLSISASDQNLASDMDTSG 599
 QY 601 LNVPRYGNGLTWGMWAKTODEPAPASATITDPOKANRPHRTLLTWLPAGVPSPKHRS 660
 DB 600 LNVPRYGNGLTWGMWAKTODEPAPASATITDPOKANRPHRTLLTWLPAGVPSPKHRS 659
 QY 661 PLIANTLMGNMLLATESLKSNAELTPSDHPFWGITGGLGMVYODPRENHPGFMHRSRG 720
 DB 660 PLIANTLMGNMLLATESLKSNAELTPSDHPFWGITGGLGMVYODPRENHPGFMHRSRG 719
 QY 721 YSAGMIAGOTHTFSLKFSOTTYTKLNERAKNNVSSKNSSCOGEMLFSLDEGLTKLVGL 780
 DB 720 YSAGMIAGOTHTFSLKFSOTTYTKLNERAKNNVSSKNSSCOGEMLFSLDEGLTKLVGL 779
 QY 781 YSAGMIAGOTHTFSLKFSOTTYTKLNERAKNNVSSKNSSCOGEMLFSLDEGLTKLVGL 840
 DB 780 YSAGMIAGOTHTFSLKFSOTTYTKLNERAKNNVSSKNSSCOGEMLFSLDEGLTKLVGL 839
 QY 841 SLSHTEVGAYPRSFSTKPLINVLPIGVKSFPMNATOPQAMVELAYQVLYRQEBG 900
 DB 840 SLSHTEVGAYPRSFSTKPLINVLPIGVKSFPMNATOPQAMVELAYQVLYRQEBG 899
 QY 901 IATQILASGIVFGSGSPSRHAMSKKISQOTPLISMLTLHFQYHGFYSSTFCNTLNGE 960
 DB 900 IATQILASGIVFGSGSPSRHAMSKKISQOTPLISMLTLHFQYHGFYSSTFCNTLNGE 959
 QY 961 IALRF 965
 DB 960 IALRF 964
 RESULT 2
 OMPF_CHLTR STANDARD: PRT: 1034 AA.
 AC P38008: 084878: 30-MAV-2000 (Rel. 30, Created)
 DT 30-MAV-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PUTATIVE OUTER MEMBRANE PROTEIN F PRECURSOR.
 GN PMPE OR C7870.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxId=813:
 OX 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/CX:
 PX MEDLINE=99000809: PubMed=97844136:

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marache R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.:
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis".
 RT Science 282:754-759(1998).
 RL [2].
 RN SEQUENCE OF 26-35.
 RP STRAIN=L2/434/80.
 RC Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,
 RA Comanducci M., Christensen G., Birkelund S., Vitreou E., Ratti G.,
 RA Pallini V.:
 RL Submitted (SEP-1994) to the SWISS-PROT data bank.
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001360; AAC68468.1; -
 DR Outer membrane; Signal.
 KW SIGNAL 1 25
 FT CHAIN 26 1034 PUTATIVE OUTER MEMBRANE PROTEIN F.
 FT CONFLICT 34 35 RR -> FH (IN REF. 2).
 FT SEQUENCE 1034 AA: 112392 MW: 445FAC35D463A67 CRC64:
 SQ
 Query Match 17.0% Score 866.5; DB 1: Length 1034;
 Best Local Similarity 28.1%; Pred. No. 76-46;
 Matches 293; Conservative 149; Mismatches 406; Indels 193; Gaps 36;
 QY 55 NCYLDNLRITLAILOKTNEGAAVITIDYLSFDFQKEGIVFAKMLPESGAI----- 108
 DB 57 NIVLSNLO-----SNGTGACTISG-----NTOIFNSVNTTLDSCGAEPMWTS 102
 QY 109 -----GYASNPSPVEIRDTGPIVFNENNTCCRFSTSNPAV-----NK 149
 DB 103 FTASDNANLFCNNCTHNKGGAIKRG-GPIRLNNDVLFYNNISAGAYVGTGDNE 161
 QY 150 IREGAIHQNLYIHNNDVYGFKNFSYKGAISTANTFVSENOGCFLEMDNICI-- 207
 DB 162 KNRGALATATTLTGNR-TLAFINNMSCDCGGAISADTOISITDVGILFENNHTLNH 220
 QY 208 -----QTNAGKGAIIYAGTS-NSFESNCDLFFINNACGAGIIFSPICSLGNRCNIVE 262
 DB 221 IDYTQENNAKRGALICSRDLDCISNSGPIYFNNGGGAISATKRVLDNNERLIF 280
 QY 263 YNNRCFKNVEYASSAASDGAIKVTRRLDVTGNRGRIFESDNTNNGGAIYAPVTVLV 322
 DB 281 SNNSLIG--WSQSSASNCGAIQTTGFTLRNNKSIYFDSNTATHAGALINCXYIDRD 338
 QY 323 NGPTFYFINNIANNKGAIIYIDGTSNKSISADRA-----IIFENYIYNTVNTANG 372
 DB 339 NCPVYFLNNSA--AMGAAP-----NLSPRSATNYIHTGTGDIVPNNNVFTLDGNLL 389
 QY 373 TSTSNAPPRNATIVASSGCEILLGAGSSONLIFYDPIE-----VSNAGS-----V 419
 DB 390 GKRLFLHINNNEITPTLS-----LGAKDRTIYFDLQWEMKENTSNPSPISRTI 445
 QY 420 SFNKEADOTGVSFSGATVNSADFHORNLOTV-----TPAPLTLISNGCLIEDHAOLTVN 474
 DB 446 IYNPETEFGAVVS--YNNQSSDIRLTMKEHNYIKEAPTLTGLTALIEDDALEIF 502
 QY 475 R--FTQT-GGVVSLGNCAVLSCKYKNGAGNSASNAITLKHIGLNLSSILKSGAEIPLLW 531
 DB 503 NIPFTQNPSTLALGSGALLTVGKHG-----KLNTNLGVLLPILIKESKSPCLIRV 554
 QY 532 EPTN-NSNNYTADT--AATFSL-----DVKLSLIDDYGNSPYESTDLTHALSSQPMLSI 583

Mon May 7 09:06:31 2001

us-09-677-752-2.rag

Page 24

[illegible]

used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis, asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.

QY 99 NTPESGGAICVYASPNSTVEIRDTIGVIFENNTCCPFTSSND-NAAV----- 147
 Db 122 -lllegfkelsfscnslavl-----paatluksgqrlttspsngltysktdllln 174
 QY 148 -----NKR-EGGALHAQNLVYNNHNDVGFEMKNSYRGGAISIRANFEVSENC 198
 Db 175 nekfsfynlsvsgdaidskltvgsiklcvfentagdgagcvvstfsmamap 234
 QY 199 FLFMDNICIOTNTAG-KGGAIA-----GTSNSFE-----SNNDLFFINACAG 243
 Db 235 lafvnnv-----agvrgglaavdgqgqvssststedpvsfmrnlaveldgnvatvg 288
 QY 244 GAIFSPICSLTGNRCNIVFYNN--RCPKN-----VETASEASDCGAI 284
 Db 289 gglys-----ygnvalflnngkclflnnvaspylaakqrlsgaasltsmnydgga 340
 QY 285 KVTTLVDVTGN-----RGRIEFDNITKNVGAIAVPVTLVDNGPTFEINNIANN 335
 Db 341 fckngagagsgnsvsgsfdegvgnvfssnvaagkggalyakklsvanccpvgfltnland 400
 QY 336 KCGAIIYIDGTSNKSISADRHAIIFENI-----VTNVTNANSTSTANPERRNAVSSS 391
 Db 401 -ggalylgsgsgslsdsdygdllfdgnlkrakenaadvngltvss-----qalsmgsg 454
 QY 392 GEILGAGSSQNLIFYDPIEVSN-----AGVSVFENKADOTGSVSGATVNSADPH 444
 Db 455 kltllrakqhqlflndplemangnnpqasakllkndgeygtgdvlfan---gsstly 511
 QY 445 QNNLQTKTAPRLTSLNGFCIEDHQLTVNRFOTGQGVVSLCNGAVALSCYKXGAGNS--- 501
 Db 512 qn-----vleggrtvlrekaklsvnsjsqvg-----slyme-agsctwf 551
 QY 502 -----ASNASTLKHGILNLSILKSGAEIPLWVEPTN--NSNNTYADTATES 549
 Db 552 vcpqpgqpaaanqltlnhlsllsllanna-----venpntppagdsphavlgst-t 606
 QY 550 LSDVKLS---LIDDYGNSPRESTD-----LTHALISQPMLSISEASDQLRSDOM 596
 Db 607 agsvltsglffiedldtdaydrydwlgnsqklnvklqlqkx--panapsdltl----- 659
 QY 597 DFGSLNVPHYGMGMLTWGMMAKQODEPAPASATITPOKANRFRHLLTLTWAGVPS 656
 Db 660 ---gmepkygygsvklaw-----dntamngpytlkatwtctgympap 701
 QY 657 KHRSPILANTMGMMLLATESLKSNAELTPSDHPFV-GITGGGLGMVYQDPRNHPGRH 715
 Db 702 ervasivpmslvgv-ldlrshalsdaavdgyscrjlvsgvsnffyhndaldqgyr 760
 QY 716 MRSSGYSAQMIACQHTFSLKFSQYTKLERYAKNNV--SSKNYSCQCEMLFSLDEGFL 773
 Db 761 ylsqgyslg---ansyfgsamfglatlevfgr-skdyvvcrsnhhacisvylstqda-- 814
 QY 774 LTLVLGLVSGDHNCHHFTYOG--ENLTSQTFERSOT-----MCGAVFPEPMKPGS 824
 Db 815 ---lcsyylfgdaftasyvgfngnmktsyctlaeesdvrtwmcnlsgelagajlrvltps 871
 QY 825 THILT--APFLGALGYSSLSHTEVGAVPFRSTKPLINVLVPIGVKGSFNMATORPO 882
 Db 872 klylneirpvggefysatdhesfleeqdgatks--ghllnlsvpygkldrcsslt-hpn 929
 QY 883 AMTVELAVCPVLYROEBGIATOLLASKGIVFGSGSPSSRRAMSKYKISQOQPLSMILHF 942
 Db 930 kysfmaaylclayrltsgtettllshpeltclafnarhgvvvrvsgmyas---llsnl 985
 QY 943 QYHG 946
 Db 986 evyg 989

AC W88422:
 XX 26-APR-1999 (first entry)
 DT Chlamydia pneumoniae surface exposed protein Omp9.
 XX Chlamydia pneumoniae surface exposed protein Omp9.
 DE Omp9: outer membrane protein 9; surface exposed protein; antigen;
 KW Infection; diagnosis; vaccine; atherosclerosis; asthma.
 XX Chlamydia pneumoniae.
 XX W09858953-A2.
 XX 30-DEC-1998.
 PD 19-JUN-1998; 98WO-DK00266.
 PF 23-JUN-1997; 97DK-0000744.
 PR (BIRK/) BIRKELUND S.
 PA (CHIRI/) CHRISTIANSEN G.
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 DR WPI: 1999-105610/09.
 DR N-PSDB: X06821.
 PT Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 PS Claim 7: Page 56-58; 115pp: English.
 XX This polypeptide comprises the novel 96.7 kDa surface exposed
 CC protein Omp9 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06821) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.
 XX Sequence 918 AA:
 SQ

Query Match 12.5%: Score 637.5; DB 20: Length 918;
 Best Local Similarity 27.2%: Pred. No. 8.7e-39;
 Matches 262; Conservative 132; Mismatches 319; Indels 251; Gaps 47;

QY 61 LRYILAILOKTPNEGAAVTIDVLSFED---TQKGIYFAKULTPESGGAICVYASPNP 116
 Db 130 lrmhaa--ptlqk9-alkitdglvlesignldne-----nasenggal-----ntk 175
 QY 117 TVEIRDTIGVIFENNTCCPFTSSNPNAVNKIREGAIHAQ-NDYINHHNDVGFEMKN 175
 Db 176 tlstgstrfvalgn-----sssggggalyaagdvlsenagjlsfgm 220
 QY 176 FSTVRGGAISTANTFYVSENGSCFLEMNDICITNTNAGKGAIIYAGTSNFEENKDLFF 235
 Db 221 satlsggalsaaenjlvismgn--liffdgckatln-----ggal----- 257
 QY 236 INNACGAGAIIFSPLCSLTGNRCNIVFYNNRCFKNVETASSEASDCGAIKVTTLVDVTGN 295

RESULT 20
 W88422
 ID W88422 standard: Protein; 918 AA.
 XX

Db 793 dafrasygfgnqhmktsytfaeesdvrdwncnclageigajpivltpsklylnelrpfv 852
Qy 834 GALTGYSSLSHFEVGAIPRPFSTKPLINLVIPGVKGSFNNATOPQAMVELAYOV 893
Db 853 qaeisvadesfleeqdqaralks-ghllnlsvpvvklfdrsst-hpkxysfmaaylcd 910
Qy 894 LYRQPGIATOLASKGIWFGSGSPSRHMSYKISQOTOPLSWLTLPFOYNG 946
Db 911 ayrtisgtetllshgwtclafharhgvvrvgsmyas-----ltsnlevy 959

RESULT 15
B13639
ID B13639 standard; Protein: 1006 AA.
XX
AC B13639:
XX
DT 02-FEB-2001 (first entry)
XX
DE C. trachomatis pmg gene protein.
XX
KM Chlamydial infection; sexually transmitted disease;
KM pelvic inflammatory disease; PID; tubal obstruction; infertility;
KM trachoma; blindness; acute respiratory tract infection;
KM atherosclerosis; coronary heart disease; antibacterial.
XX
OS Chlamydia trachomatis.
XX
PN MO200034483-A2.
PD 15-JUN-2000.
XX
PF 08-DEC-1999; 99WO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
PA (CORI-) CORIXA CORP.
PI Probst P, Bhattacharya A, Skeiky YAM, Elling SP, Jen S, Stromberg EJ:
DR WPI: 2000-431303/37.
PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
PT comprises immunogenic portion of Chlamydia antigen, which comprises
PT amino acid sequence encoded by polynucleotide sequence -
XX
XX
PS Claim 2: Pages 208-210; 256pp; English.
XX
CC The present invention relates to new nucleic acid sequences and the
CC proteins encoded by the nucleic acid sequences. The encoded proteins
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
CC proteins are useful for the serodiagnosis and treatment of Chlamydia
CC infection. Chlamydiae are intracellular bacterial pathogens that are
CC responsible for a wide variety of human infections. C. trachomatis
CC infection is one of the most common sexually transmitted diseases and can
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
CC and infertility. Trachoma due to ocular infection with C. trachomatis is
CC the leading cause of preventable blindness worldwide. C. pneumoniae is a
CC major cause of acute respiratory tract infections in humans and is also
CC thought to play a role in the pathogenesis of atherosclerosis and
CC coronary heart disease. The present sequence is a protein isolated in the
CC present invention.

SQ Sequence 1006 AA.

Query Match 12.7%; Score 647; DB 21; Length 1006;
Best Local Similarity 25.4%; Pred. No. 2e-39;
Matches 257; Conservative 156; Mismatches 388; Indels 212; Gaps 43;

Qy 39 SLNKKISLTGDTHTNFCYLDNLKYLALLOKTPNEGAATITDYLSEFFDQKGIYFAK 98
Db 78 nlgsfvlvgqshlt-----fenlr-----tslngals-----nsaagdlf--- 115
Qy 99 NITPESGALCYASNSPTVEIRDITGVIFENNCCAPFFSSNP-NAAV----- 147
Db 116 -tlegfksltsncslavyl-----paatlkyagqrltcltscpsngtlskcdllln 168
Qy 148 -----NKIR-EGGAIHAQNLININHDVGVKFNESYVRGGAISTANFVVENOS 198
Db 169 .nekfsfynlvsqdgaldakslvqglsklcvfgentagadagacqvlsfsameap 228
Qy 199 FLFMDNICIOTNTAG-KCGAIYA-----GTSNFE-----SNCDLFFINNACAG 243
Db 229 lafvanv-----agvrggilaavdgqgvssstledpvsfarnlavefdgnvavrg 282
Qy 244 GAIFSPICSLTNRGNIFVYNN--RCFKN-----VETASEASDGGAI 284
Db 283 ggiys-----ygnvalflmgkclflnvvaspyiaakqrlsqasntsnnygdgga 334
Qy 285 KVTTRLDVTGN-----RCRIFESDNITKNCGAIYAPVTLVDNGPTVFINNAN 335
Db 335 fckngaqgsmnsgsvsldgegvvlfssnvaagkggalyaklsvancgqvqlnland 394
Qy 336 KCGAIYIDGTSNSKISADRAHIIFENDI-----VTNVTNANGTSTANPPRRRAITVA 391
Db 395 .ggailylgesgeislisdgydlldgnlkrakenaadvnglvss-----qalsmgs 448
Qy 392 GEILLGAGSSONLIFYDPLEVSN-----AGVSVPFNKADOTGSVFSGATVSA 444
Db 449 klitlrakagbhlfrndpiemangmpqssklklkindgeytdvifan---gsstly 505
Qy 445 ORNLQTPAPRLTSLNGFLCIEDHAQLTVNRTQGVVSLGNGAVLSCYKNGAGNS--A 502
Db 506 gn-----vlieqgrivlrekeklsvnsisqgslymeagstldfvlrppqp 556
Qy 503 SNASTITLKHIIGLSSILKSGAEIPLWVEPN--NSNNTADTATPSLSYKIS--L 557
Db 557 anqltlnlshlssllanna-----vlnpplnppagdhavlsst-lagsvltsp 611
Qy 558 IDYGNSPYESTD-----LTHLSSOPMLSTSEADNQLRSDMDFGLVNPHYG 607
Db 612 feolodctayrlyowisngklnvklqigtkp--panapsdltl-----gnempky 661
Qy 608 WQMLTWGMAKTODPEPASSATITDPOKANRPHRTLLTLPAGVVPSPKHSPIANTL 667
Db 662 yqgswklaw-----dntanngpytlkalwtklypppervaslvpna 706
Qy 668 WGNMLLATESLKNASALTPSDHPF--GITGGGIGMMVYODDPRENNHGFMRSSGSACMI 726
Db 707 wgs-lldirsahsaibaasvdgrsyrcrglwgsvsnfilyrdalaggyryisgy 763
Qy 727 AGOTHTFSLFSQTYTKLNERAKNNV--SKNYSCQGMFLSLDGFLLTKLVGLSYG 784
Db 764 -ansfygsmfgaifevgr-skdyvcrsnhaeigvyslstgq-----lcsy 816
Qy 785 DHNCMHFYTG-ENLTSQGTFRSQT-----MGAVFEDLPMPKFGSTHILT--APL 833
Db 817 dafrasygfgnqhmktsytfaeesdvrdwncnclageigajpivltpsklylnelrpfv 876
Qy 834 GALTGYSSLSHFEVGAIPRPFSTKPLINLVIPGVKGSFNNATOPQAMVELAYOV 893
Db 877 qaeisvadesfleeqdqaralks-ghllnlsvpvvklfdrsst-hpkxysfmaaylcd 934
Qy 894 LYRQPGIATOLASKGIWFGSGSPSRHMSYKISQOTOPLSWLTLPFOYNG 946
Db 935 ayrtisgtetllshgwtclafharhgvvrvgsmyas-----ltsnlevy 983

RESULT 16
Y16737
ID Y16737 standard; Protein: 1013 AA.
XX

```

Db      455 gklttlrekagbqllfndplemangnqpaseepikindgeytgdlvf--angnstly 512
OY      444 HORNLQTKTPAPLTLNSGFLCIEDHAQLTNRFTOTGGVNSLGCAGVLSCKKNAGNS-- 501
Db      513 qn-----vlieqgrlvlfrekaklsvnslsqglsglymeagstldvltcpqpqp 562
OY      502 ASNASITLKHIGLNSLSLKSGAEPLPLWEPN--NSNNYTAADTAATFSLDVKLS--- 556
Db      563 aangeitltsnlhlstslsllanna-----vtnptnppagdsnpavlgst-tagpvtisgpf 617
OY      557 LIDDYGNSPYSTD-----LTHALSSQPMLSISEASDNLRSDDMDFSGLNPNHY 606
Db      618 lfedldtdlaydrydvgnsqkldvklqlglqfp--saanapadlrl-----gnempky 667
OY      607 GMOGLMTGMMAKTQDPERASSAATITDPOKANRPHRTLLTWLPAGYVSPKHRSPLIANT 666
Db      668 gyggswwklaw-----dptlanapylkalwcklqyngpgervaslvpns 712
OY      667 LMGNNLLATESLKNASAEPLPSDHPFW-GITGGGLGMVYQDPRNHPGFIHRRSGYSAGM 725
Db      713 lwgs-ldlrshaaigavsgvgrcgylwsgvsnfgyhdrdaigqgyrlysggsylg- 770
OY      726 TAGOTHTFSLKFSQTYTKLNERYAKNNV--SSKNYSQCGEMLFSIQEGFLLTKLVGLTSY 783
Db      771 --ansyfgssmfglafllevfgr-ekdyvvcrsnhacisgsvylstkqa-----lsgsylvf 822
OY      784 GDHNCNHRYTQG-ENLTSQGTFRSQT-----MGCAVFPDLPKPKPGSHILIT--APF 832
Db      823 gdafllrasygflngqumktsyltaeesdvrdnncvgeigvylpvtlpsklylnelpf 882
OY      833 LGALGATYSLSHFTFEGVAGYPRSFSTKPLINVLNVLIGVKGSEFMATQPOAVTELAYOP 892
Db      883 vgaeisysdhesfteegqatarafs-qhlmlslvpygvkfdrcsst-hpnkysfmgaylc 940
OY      893 VLYROEPIATQLASKGIWFGSGSPSSRHANYSKISQOTOPLSLTLHFOYHG 946
Db      941 deryctisgqtltlshgclwtldaftharhylvrgsmas----ltanleyvg 990

RESULT 13
Y69369 18-SEP-1998; 980US-0097187.
ID      Y69369 standard; Protein; 918 AA.
XX      XX
AC      Y69369.
XX      XX
DT      19-JUN-2000 (first entry)
DE      Amino acid sequence of the CPN100395 polypeptide.
XX      XX
KW      CPN100395; Chlamydia infection; immune response; vaccine.
OS      Chlamydia pneumoniae.
XX      XX
PN      WO200011183-A2.
PD      02-MAR-2000.
XX      XX
PE      18-AUG-1999; 99WO-1B01449.
PR      20-AUG-1998; 980US-0097187.
PR      20-AUG-1998; 980US-0097188.
PR      20-AUG-1998; 980US-0097189.
PR      20-AUG-1998; 980US-0097190.
PR      20-AUG-1998; 980US-0097195.
PR      20-AUG-1998; 980US-0097196.
PR      20-AUG-1998; 980US-0097197.
PR      27-AUG-1998; 980US-0097191.
PR      17-AUG-1999; 99US-0376770.
XX      XX
XX      (CONN-) CONNAUGHT LAB LTD.
PI      Mordin AD, Oomen RP.

```

```

XX      XX
DR      WPI: 2000-224703/19.
DR      N-PSDB: 261509.
XX      XX
PT      Novel antigens and corresponding DNA molecules that can be used to
PT      prevent, treat and diagnose disease caused by Chlamydia infection in
PT      mammals, especially humans -
XX      XX
PS      Claim 19; Fig 15-E; 201pp; English.
XX      XX
CC      Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
CC      are present in the bacterial membrane structure, in the external
CC      vicinity of the membrane structure, in the inclusion membrane structure,
CC      structure, in the external vicinity of the inclusion membrane structure,
CC      and in the cytoplasm of the infected cell. The polypeptides may be
CC      used to prevent, treat and detect the presence of Chlamydia infection
CC      and/or the presence of Chlamydia in a sample. The polypeptides may
CC      also be used to induce an immune response in a mammal. The vaccine
CC      vector comprising the polynucleotides is used to induce an immune
CC      response in a mammal. Antibodies directed against the polypeptides
CC      may also be used therapeutically to treat and/or prevent a Chlamydia
CC      infection.
XX      XX
SO      Sequence 918 AA:

Query Match      12.8%; Score 649.5; DB 21; Length 918;
Best Local Similarity 27.2%; Pred. No. 1,1e-39;
Matches 261; Conservative 134; Mismatches 322; Indels 243; Gaps 46;

OY      61 LRYTALIQKTPNCGAATITDYLSFPTQKEGIFAKNLTPESGALIGVSPSPYEI 120
Db      130 lrmiaa-prltgy-alkldgvlf--esigndlneassenggal-----tkltsl 179
OY      121 RDTIGPVTFENNTCCRPFTSSNPNAAVNKIREGAIHAO-ULYINHNHDVGFPMKFSYV 179
Db      180 lgstlrfvalfgn-----ssqggalyasgdsvalsenagilsfmsaett 224
OY      180 RCGAISTANTFVSENOGFLFMDNICIOTNPAKGAIVAGTNSFESNCDLFFINNA 239
Db      225 sgaisaegnlvisnqn--lffdgckatn-----gga 258
OY      240 CCAGATFSPICSLTGANGNIVFYNNRCFKNVERTASSASOGCAIKVTRLDVGNRGRI 299
Db      259 ckgagapndpiltlsgne-sihflnn-----tagnggal-ykkkvlvssqrgv 306
OY      300 FFSQNTKNNCGAIVAPVTLVDNCPYFINNIANNKGAIVYIDTSSKTSADHAIIF 359
Db      307 lfsnkaan-----alpkgaalaldsgelsisadgnl 342
OY      360 NENIVTNANGTSTSANPP--RRNAITVASSGCEILLGAGSSONLIFYDPIEVSNAV 417
Db      343 -----egntstlsgspasvtrnaidlaenakflnlratrgkvlyfipitssgtd 393
OY      418 SVSEKKEAD-----QTGSVVFSGATVNSADPHQ-RNLOTKTPAULTLSNGFLCEDHAQ 470
Db      394 klslnk-adagsgnltyegylvifsgelkseeejlkpndnlkstftgaveiaaagvlkdqvt 452
OY      471 LTVNRFOTGCGVNSLGNAGVLSGCKYKNGAGNSASNSITLKHIGLNSLISLGAEPILW 530
Db      453 vvanliltlyegskvymdg-----gltfesaagvclngalhlidl 493
OY      531 VEPITNNSNNYTAADTAATESLSQVLS---LIDDYGNSPYSTDLTTHALSSO---PMLSI 583
Db      494 -dgtnk-----alikataaskdvalsgpmlvdagyny-yehe---hnlsgqvvyfll 542
OY      584 SEASDNLRSDDMDFSGLNVP-HYGMQGLMTGMMAKTQDPERASSAATITDPOKANRPHRT 642
Db      543 s-agqtmctldipdpllnthhygyqgnwlvw-----datakt-----kn 585
OY      643 LTLTWLPAGYVSPKHRSPLIANTLMGN-----MLLATESLKNASAEPLPSDHPFW 692
Db      586 altlwtktgkxpnperqglvlnpslwgstvdvrsiqslmdrstslsstlnlwg----- 640

```

PR 29-OCT-1998: 98US-0106074.
 PR 29-OCT-1998: 98US-0106087.
 PR 02-NOV-1998: 98US-0106587.
 PR 02-NOV-1998: 98US-0106588.
 PR 02-NOV-1998: 98US-0107034.
 PR 02-NOV-1998: 98US-0107035.
 XX (CON-) CONNAUGHT LAB LTD.
 PI Murdin AD, Oomen RP, Wang J;
 DR WPI: 2000-350688/30.
 DR N-PSDB: A28708, A28709.
 XX
 PT Chlamydia antigens and the proteins they encode, useful for
 PT vaccinating against Chlamydia infections that affect the respiratory
 PT tract
 PS
 PS Claim 13: Fig 19; 226pp; English.
 XX
 CC The nucleic acids may be used for the recombinant production of the
 CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
 CC recombinant DNA methodologies. The polypeptides may then be used to
 CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
 CC C. pneumoniae, are pathogens responsible for upper respiratory tract
 CC infections such as community acquired pneumonia, acute respiratory
 CC disease and bronchitis and may be implicated in atherosclerotic changes
 CC and asthma. The nucleic acids may also be used as probes for detecting
 CC the presence of Chlamydia nucleic acids in samples (and therefore
 CC diagnose infections) and the proteins may be used as antigens for the
 CC production of antibodies that may be used to detect Chlamydia proteins
 CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
 CC
 SO Sequence 788 AA:

Query Match 18.0%; Score 915; DB 21; Length 788;
 Best Local Similarity 30.5%; Pred. No. 1.5e-59;
 Matches 255; Conservative 143; Mismatches 365; Indels 74; Gaps 27;

QY 153 GCAT-HAONLTYNNHNDVGFPMKNESYVVGCAISTANTFVSENOISCFPMNICIOTNT 211
 DB 2 9GAVLSCKSVLTKNGLEqYfInnkakssggaIgaallInkntqclff-----nna 54
 QY 212 AC--KGGAIVACTSNSESNMCDLFEINNACCAGAI-FSPICSLTGNRGNTIVFVNNRCF 268
 DB 55 aqgtacgaIfanaCr-IemnsqpiYfInnqsgIgaItrhgceIlktkngsvifInnn--f 111
 QY 269 KAVETASSEASDGAIKVTRLDVYGNRGRIFFSDNITKNYGAIYAPVTVLVNCPYF 328
 DB 112 ameadtsanhsaggaIyCis-csIkdnpgjaafdmntaardgaIctqsltdqsgpyvf 170
 QY 329 INNINANKGAIYIDCTSNKISADRAHIIFFENIVTNTNANGSTSNAPRRARITVA 388
 DB 171 Inn-9qtwggaImIrgdgacclIfadqgdllfYnnrhfkclftsn--hvsYnctnvsIvlg 227
 QY 389 SSSGELLGAGSSONLIFFYDPI-----EVSNAGVSVPFNKEADOTGSVVFSG-----ATVN 439
 DB 228 aseq-----hsatfYdpIlIqYtIqns--IqkfnpenelgIltIsayIpdtsIs 276
 QY 440 SADF--HQRNLOTKPAPILTSNGFLCIEDHQAOLTVNRFTQTGGVVSILNGAVLSCKYNG 497
 DB 277 rddfishfrn-----hIglyngIaliedraewkykfIdqfgtlrlgsravfstIdga 329
 QY 498 AGNSASNAISITLKHGLNLSILKSAEIPILMWEPTNNSNNTATPATFISLSDVKLSI 557
 DB 330 qsssvsvvInnInnaInIpsIl-gnrvaPkIwIrlptgsaPySednplInIs9-plsI 387
 QY 558 IDYGVSPYESTDLTALISQPMISISEASDNQLRSDDMFSLN-VPHYGMGLTWGCG 616
 DB 388 IdenIdpYdtdadIagIeapIlylIdvtakhInIdnfYpegInltqpygygvsapYw 447
 QY 617 ANTQDEPAPASATITDQKANRFHRTLLTLWLPAGYVSPFKHRSPLIANTLW--GNMLL 673

DB 448 IeIttdts-----edvnlhrqIygdwCpCykvnpenkIdIsafwsgfhnIa 502
 QY 674 ATESIKNSAEITPSDHPWFGITGGCIGMMVYQDPRNHPGFHMSSGSGSAGMIA--COTH 731
 DB 503 tIrytqgqIapt-----asgeatrlfYnqsnndakgIhmaeIqsgIgtntsnh 556
 QY 732 TFSLKFSQTYTKLNERVAKNNYSKNYSCQEMLEFS-LOEGFLLTKLVGLSYGDHNCIH 790
 DB 557 sfynfsgIflsnIyesnIsdsvahtttvaIqInnpwIqfscsIsaIa-ysyn---hh 612
 QY 791 FYTOGEN--LTSQGTFSQTMGAVFFDLPKRPFGSHILAPLALGIRYSSLSHTEV 848
 DB 613 IkasysgskIqtekgcyStIlgaaIsclsIq-wrsrIpIhtpIltIqIavrsnqIatqes 671
 QY 849 GAYPFSFTKTPILNVIAPIGVKSGFMNATORPOAWTELAYOPVLYROEFGIATOLLAS 908
 DB 672 gdkarkIsvhkpIynIvlpIqIgaeswkfrIplynIeIaypvlYqnpelInvsIess 731
 QY 909 KGWFGSGSPSSRHMSYKISQOTQPLSLTLHFQYMGFYSSSTFCNYLNGEIALRF 965
 DB 732 gswllsgstIarnaIatfkgrnqIIfpklsvIldyqgsvssstthylhagttfkf 788
 RESULT 11
 ID Y35088 standard; Protein; 822 AA.
 AC Y35088;
 DT 13-SEP-1999 (first entry)
 XX Chlamydia pneumoniae surface exposed polypeptide.
 DE
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-1B01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R.
 XX
 DR WPI: 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS Page 971-973; Disclosure: 1912pp; English.
 XX
 CC Y34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see X01990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see Y34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 SO Sequence 822 AA:

XX CPN100624 RX-64: antigen; infection; diagnosis; therapy; vaccine;
 KM outer membrane protein.
 XX Chlamydia pneumoniae.
 OS WO200039158-A1.
 XX 06-JUL-2000.
 PD 23-DEC-1999: 99WO-CA01230.
 XX 23-DEC-1998: 98US-0113280.
 PR 23-DEC-1998: 98US-0113281.
 PR 23-DEC-1998: 98US-0113282.
 PR 23-DEC-1998: 98US-0113283.
 PR 23-DEC-1998: 98US-0113284.
 PR 23-DEC-1998: 98US-0113285.
 PR 23-DEC-1998: 98US-0113385.
 PR 28-DEC-1998: 98US-0114050.
 PR 28-DEC-1998: 98US-0114056.
 PR 28-DEC-1998: 98US-0114057.
 PR 28-DEC-1998: 98US-0114058.
 PR 28-DEC-1998: 98US-0114059.
 PR 28-DEC-1998: 98US-0114061.
 XX (CONN-) CONNAUGHT LAB LTD.
 PA Murdin AD, Oomen RP, Wang J;
 PJ WPI: 2000-452369/39.
 XX N-PSDB: A50038.
 DR Novel Chlamydia polynucleotides and polypeptides useful for diagnosis,
 PT prevention and treatment of Chlamydia infection in mammals -
 XX Claim 16(a): Page 76-80: 215pp: English.
 XX The present sequence is that of Chlamydia pneumoniae antigenic
 CC protein CPN100624 RX-64, a putative outer membrane protein.
 CC It is an example of C. pneumoniae antigenic polypeptides of the
 CC invention (see Y95543-55) that are encoded by open reading frames
 CC (see A50030-42) identified in the C. pneumoniae genome. The
 CC polypeptides are useful in the diagnosis, treatment and prevention
 CC of Chlamydia infection. They can be prepared by recombinant
 CC methods using transformed unicellular host cells. The
 CC polypeptides, or immunogenic fragments of them, or fusion proteins
 CC that include an additional polypeptide such as a heterologous
 CC signal peptide or a polypeptide having adjuvant activity, are
 CC utilised as vaccines to protect against Chlamydia infection.
 CC The polypeptides are also used in diagnostic kits and in methods
 CC of detecting Chlamydia infection.
 CC
 XX Sequence 963 AA:
 SO

Query Match 19.5% Score 989.5; DB 21; Length 963;
 Best Local Similarity 30.0% Pred. No. 5,9e-65;
 Matches 304: Conservative 146; Mismatches 416; Indels 149; Gaps 34;
 9 LGNSISGLA-----REVSRIFLMP---NSVPPTKESLSNKLISLADGTINLT----- 54
 30 lsgnevpnlascgmsrksdtsafhtspsfriinvleplvasfrpsnlnfgndltqdtli 89
 55 -----NCIYD-NLRKYLALLOKTPNEGAAVTTIDYLSFPDIOKEGIVFAKKNLTPESGAI 108
 90 tgnslinsvidnyhy-----edqgilaacknl--fisenkgnlsferenshs99gai 138
 109 GVASPNSPVEIRDTIGPVIFFENNTCCRPFTSSNPN-----AAVNKIREGGAIA 158
 139 -----ysvre-----C--wiknqnyfslsnaaslatltsgf99gai 175
 159 ONLIINHNDVVGFKNFSGYVRGGAISTANTFVSENOCSLFMDNICIQITNTAGKGA 218
 Db 176 ldsyltnlmgqgflidhvsknrgaiyvgvslstldhlpivlkngltedsf--99gi 233
 Qy 219 YAGTSNSFESNNCDLFFNNMCCAGGAIFSPICSLTGRCGVIVYNNRCFNVETASSFA 278
 Db 234 fcravn-lernyqlqldnasggvvyflpilylsnkeileismsaasintasaqkl 292
 Qy 279 SDGAIAKVTTRLDVTGN-RGRIFESDNTKNYGAIV-----APVVLVDNGPTTFINNI 332
 Db 293 ypgggymctslshempkylif--nmtaalsggyrdrldssakiv-----traflns 346
 Qy 333 ANNKGAIYIDG---TNSKISADRHAIIFENIVTNVNTANCTSTANPERRNATIVA 388
 Db 347 atsggalinlsglstcpnfeflsadqgllfnmtli-----tsspggyrinal-ya 397
 Qy 389 SSSGFTLLGAGSSONLFFYDPFEVSNAGVS--VSPNKEADOTGVSFVGATVNSADPHORN 447
 Db 398 apglnlkigatqgyklllydpidndqtltdpivlnyephnlqvlvlsqindvsnatpnl 457
 Qy 448 LOTTPAPRLTNSGFLCIEDHAQOLTVNRFTOTGCVSLGNGAVLSCKYKNGASNASNI 507
 Db 458 flskfnsrrierylaidraalsckltsqtgllrignaalrl--ckpg-----ssi 510
 Qy 508 TLKNIGLWSSILKSGAEIPLWEPYNNNTADTATSLSDVKSLIDDYGNSEYE 567
 Db 511 nfaialnlpsllqseasapkfwiypiltglsyedsstslstllt-pitfindenepyd 569
 Qy 568 STDPLTHALSSQPMLSISEASONOLRSDMDPFGSLNV-----PHYGMQGLTWGAKROD 621
 Db 570 slldl-----seprkdiplpplpprcdckkldtsnlvleamnldehygyqslwpywmet-- 622
 Qy 622 PEPAASATITDPOKANREHRTLLTLWLPAGYVSPKHSPLIANTLV-----GNML 673
 Db 623 ---ttstsvpeqnltnhrqlyvdwtpvgyrperhgefiantlwgaynallqirll 679
 Qy 674 ATESLKN---SNELTSPDSHPFWGITGGGLGMVAYDDPRENHPGFMRSSGSAGIA--G 728
 Db 680 ppqnlkndleasliq-----glllndhnrqtkyfrnhtlyaatlsakta 727
 Qy 729 QHTPFLKFSQTYTKLNERVYAKNNVSKNYSCQGMFLSLQGFLLTKLYGYSYGDNN- 787
 Db 728 arhsfslftaqmfsktrcrqspstssmlyagllfddslldfslslysgygdhnm 787
 Qy 788 -CHHFYTOGENLTSGTFRSOTMGAVFFDLPKPEGSTHL-TAPFLGALGIYSSLSHF 845
 Db 788 lch--ytelkqskafnbnhlvasl--dctflparitrltelqflsajalrcsqsf 843
 Qy 846 TEVGAVPNSFSTKTLINLVPIGVKSGFMNATORPQAMTVELAQPVLYRQEPQIATOL 905
 Db 844 qetghlhlkflhpkhplldlsspdigrsewkshhlpmlwtelssvpltyknpmpftll 903
 Qy 906 LASKGIWFGSGSPSSRHAMSYKISOOTOPISWLTLLFQYHGYSSTFCNYLNGE 960
 Db 904 llsngltwtlqatpvyvsvaaklntsqflsrlvllsldysqyvsstvygylkae 958
 RESULT 9
 Y92830
 ID Y92830 standard: Protein: 954 AA.
 XX Y92830:
 DT 29-AUG-2000 (first entry)
 DE C: pneumoniae CPN100624 antigen.
 XX Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
 KM anti-arteriosclerotic; vaccine.
 OS Chlamydia pneumoniae.
 XX WO200024765-A2.
 PN

```
Db 616 klllyanwsplyrphpergetflnalwgsayalaglnhiss-----wdeekyha 667
Oy 698 ----GLCMWYODPRENHPCGHMRSSGYSAGMIA--GOTHTFSLKFSQTTTKLNERAKN 751
Db 668 aslqglglvqkdkngfkgrshmlgysaltcaatsqpsrlqfqlfiskakehesqn 727
Oy 752 NVSSKNV---SCGEMFLPSLOEGFLTLTKVLG-LYSGDHNCNHFYTOGENTISQGTFRSQ 807
Db 728 stshhyfsgmientlf--kewirls--vsajamfsehnhtmyqgllegnsgsfmnh 783
Oy 808 TMGCAVFFDLPMKPFGSTHILTAFLGALGYSSLSHFTVGAVPRSEFTKPLINLVLP 867
Db 784 llaagalscvflpghges-lqlypfitalaigrnlaaidesgdharefslhrpltdvalp 842
Oy 868 IGVGSEPMATQROQAWTVELAYQPVLYROBPQIATQLLASKGIWFGSGSPSSRHAMSYK 927
Db 843 vglraswkhhrvplvltelsyrsfilyrqdbelshklllsqglwtlqatrpvtynalgik 902
Oy 928 ISQOTPLSMTLHFQYHGFFSSSTFCNYLNGEIALRF 965
Db 903 vktcmqvfpkvtlslsdyssadlssstslshylnvasrmrf 940

RESULT 6
Y35083
ID Y35083 standard; Protein; 969 AA.
AC Y35083;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed polypeptide.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KV vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN MO9927105-A2.
XX
PD 03-JUN-1999.
XX
PE 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
PA (GEST ) GENSET.
XX
PI Griffiths R;
XX
PI WPI; 1999-357842/30.
XX
DR Genome sequence of Chlamydia pneumoniae
XX
PS Page 967-969; Disclosure; 1912pp; English.
XX
CC Y3584-Y3879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see X91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see Y34584-Y3879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 969 AA:
```

```
Query Match 20.8%; Score 1057.5; DB 20; Length 969;
Best Local Similarity 30.5%; Pred. No. 5,2e-70;
Matches 302; Conservative 165; Mismatches 447; Indels 77; Gaps 28;

Oy 6 FFELIGNSLGLAREVDSRIFLMNSVDP-----PKRESLSNK-----ISL 46
Db 25 flfylygmiaemfcmtpavyslqdslekfalerdeefrsfpildslstlgtfspiltf 84
Oy 47 TGDTHNLT-NCYLDNLARILIAILOKTPREGAAVTITQVLSFDTQKGIYPAKULPRESG 105
Db 85 vgnrhnsqgdvlvnsyxdnlllwtsagavscnml--lsmvedafalskllaigtq 142
Oy 106 GAICYASPNSPTEVIRDTIGPVIFENNTCCRPFTSSNPNAAVNRKREGAITHAO-NLYIN 164
Db 143 galacqg----actlknrgpliffnrgln-----nastggetrggalacngdtlis 191
Oy 165 HHNDVGMKMFSTYVRGCAISTANTFVYSENOCFLEMDNICIOTNAGKGAITYAGTSN 224
Db 192 qnqglfyfvnsvnmwggalslmgchcrfqnrapliffn-----nlaps99galrsent 245
Oy 225 SPESNNOCLFEFINNACGAGALFSPICSLTGNRGNIVFYNNRCFKNYETASSEASDGA 283
Db 246 ltsdnlrplylknmcgmnggalqtsvalknsgsvlfnnmlalsg--slngsgs9ga 303
Oy 284 IKVTRLDVTGNRGRIFESDNTITKNYGAIYAPVTVLVNDNGPTYFINNANKGAIYID 343
Db 304 l-yltnlsiddnpgtllfnnmycitrddgaictqflltknshvyftnn-qgnwggalml 361
Oy 344 GTSNSKISADKHAITFNNIYTNWTNANGTSTSANPPRNMTAASSGELLGAGSSQN 403
Db 362 qdstclllaegnalafqnevflltfg-----lynalhcfnps-nlqganqyrt 410
Oy 404 LIFVDPTEVSNAGVS-VFENKEADQTSVVFSGATVNSADRHORLQKTPAPLTLNSGF 462
Db 411 tafidplehpbctnpllfnnpnaingtllfssayipaasyennffissknteeingv 470
Oy 463 LCIEDHAQLTVNRFTQGTGGVYSLGNAVALSCYKNGAGNSAS-MSITLKHIGLNLSTLK 521
Db 471 lstedragwqfykftqkgjllkghaaslattnseptsavsgvllnnlalnlpisa 530
Oy 522 SCAETPLLWEPPTNNSNYTADTATFSLSDVKLSLIDYGNSPRESLDLHALSSQML 581
Db 531 kg-kapllwlrplqgsapfitednpltlfsg-plllneendpydsldseplqnlll 588
Oy 582 SISEASDNQLRSDMDVFGSLN-VPHYGQGLMTWGMARTQDEPPASSATITDPOKANRF 640
Db 589 slsdvtarhndlnhpeslnatehygyglwspyvet--littnnasi---etancily 643
Oy 641 RTLLTLWLPAGVPSPKHRSPLIANTLWGNMLLATESLKSNAELTPSD--HPFWGITGGG 698
Db 644 ralyanwprlygkknvpygdlatplwgsfhtmtslrsyrtgdslderplelqgl 703
Oy 699 LGMWYODPRENHPCGHMRSSGYS--AGMIAGOTTFSLKSQTTTKLNERAKNANVSK 756
Db 704 dglfwhqnsipapqfrlqstlqyslqasasetlshklsldgicftrtckelqssnmvsaah 763
Oy 757 N--YSCGEMFLPSLOEGFLTLTKVLGYSGYGDHNCNHFYTOGENTISQGTFRSQGAVF 814
Db 764 ntvsalylwelp--fgeaf-atsvtlayygdhlnslhpsbqe-qaegtvyshltaaal 820
Oy 815 FDLPMKPFGSTHILTAFLGALGYSSLSHFTVGAVPRSEFTKPLINLVLPICVKGSF 874
Db 821 csfpqgkeyhl--spfgatairshqgateefeeikfvsqpfylyltpilqgkw 878
Oy 875 MATQROQAWTVELAYQPVLYROBPQIATQLLASGIGWFGSGSPSSRHAMSKKISQGTQ 934
Db 879 gskflwprlewletlsyeplygqnpqlyvllaasgswdlghnyvrnalgykvhmqal 938
Oy 935 LSMVLTHFQYHGFFSSSTFCNYLNGEIALRF 965
Db 939 frsldflidygsvssstethlqagstlkf 969
```

OY	136	RPFSSNNNAVNRKREGGALHAQMLYNNHHDDVYGEFKNFSSYRGGAISTANFVSEN	195
Db	149	rlfctwrnyaa-dkrtregga lhaqnl ylnmhddvqfinkfssyvgga lscanctfvsen	207
OY	196	QSCFLFMDNICIQOTWTAGKGAIVAGTSNSFESNNCDLFFINNAACAGAI FSPICSLTG	255
Db	208	qscflfmdhclclqntlaqkga lyaqtsnfesnmcdlffinnaccagailfspicsltg	267
OY	256	NRGNVTFNNNGCFKVEFLAASSAASGGCAIKVTTRLDVTGNNGRIFFSQNIIRKNNGCAIYA	315
Db	268	nrgnvlvfnmncfknvelaassaaodgga lkvctrltdvgngrlffsdh ltknyga lya	327
OY	316	PVVTLYVDNGPTFYFINNINNNKGAIYIDGTSNSKISADRAHAIIFENENVTIVNTNANGTST	375
Db	328	pvtvlvdngpctfyfnninnnkgaa lya ldgtsnksaadraha lfnenlvtnvtnangtst	387
OY	376	SANPRRNAITVAASSGGEILLGAGSSQNLIFYDPIEVSNAQSVSFNKADDTGSVFSG	435
Db	388	sanprrna ltvassggelllgagssgnllfydpilevsnagvsvfnkadtgsvfsq	447
OY	436	ATVNSADRHQRLQTKTAPRPLTSLNSGFLCIEDHADLTNNRTPOTCGVSLGAGVALSGYK	495
Db	448	atvnsadrhqrlqtktpaprlt lnsqflciedhaqltvnrftqgvsvslgagvalsgyk	507
OY	496	NGAGSASNASITLKHIGLNLSSILKSAEAIPLWVEPTNNSNNNTADTAATFSLSDKL	555
Db	508	ngtgsaasnas ltkhlglnlsslksgaeapl lwpvptnnsnyttadtatfslsdvkl	567
OY	556	SLIDTGYNSPRTSTDLTAHASSQPHLSTSEADNDLBRDDMPGSLNYPHOGMOGLWIMG	615
Db	568	slldtgyngsprstldtahnassqphlsts eadndlgse nldfsg lnpvhywqglwtwq	627
OY	616	NAKTDOPPEASSATTTDPOKARFRTLLITWLPAGVYSPSPKRSRPLANTLWGMMLLAT	675
Db	628	waktodppeasatttdpokarfrrtllitwlpagvy spsprksrplantlwgmml lat	687
OY	676	ESLKNSAEITPSDHPFWGTTGGGLGMNAYODPRENHFGFHMSSGYSAGMIMGOTHTFSL	735
Db	688	eslknsaelt psdhpfwg ttggglgmnyodprenhpgf hmtsssgysagmtimgothtfsl	747
OY	736	KESQNYTKINEEYAKNNVSSKNVSCQGEMLFSLQEGFLTLTKLVGLYSYGDHNCHEFTYQOQ	795
Db	748	ktsqnytkineyaknnvssknyscqge m lslqeg fl ltklvgl ysygdhnc hieft yqoq	807
OY	796	ENLTSQGTFRSOTMGAAVFEDLPMKRPGSTHLLTAPFLCAGICITYSLSHFTTEVGAIPRSF	855
Db	808	enltsqgtfrsqtmgavf edlpmkrpgst hllt apflcag lcslyslshfttevgayprsf	867
OY	856	SRKTPILNVLIVIGKSGSMNNAFORQAAFTVELAOPVLYRDEPGATOLLSKGIWEGS	915
Db	868	s rktplnvlivlvgksgsmnnaforqaaf tvela opvlyrdepgatol llskgiw egs	927
OY	916	GSPSSRHMASTYKISQOTPLSNLTLTFOYHGCYSSSTFCFNVIINGEIALRFP	965
Db	928	gspssrhmastrykisoqtplsnl tltfoyhgcysstfcfnv iing eialrfp	977
RESULT	4		
Y92833			
ID	Y92833	standard; Protein: 931 AA.	
XX			
AC	Y92833:		
XX			
DT	29-AUG-2000	(first entry)	
XX			
DE	C. pneumoniae CPN10628 antigen.		
KW	Antigen: anti-inflammatory; respiratory; antibacterial; anti-asthmatic;		
XX	anti-arteriosclerotic; vaccine.		
OS	Chlamydia pneumoniae.		
XX			
PN	WO2000024765-A2.		

```

XX 04-MAY-2000.
PD
XX
PF 28-OCT-1999; 99WO-CA00992.
XX
PR 28-OCT-1998; 98US-0106034.
XX
PR 28-OCT-1998; 98US-0106039.
PR 28-OCT-1998; 98US-0106042.
PR 28-OCT-1998; 98US-0106044.
PR 29-OCT-1998; 98US-0106072.
PR 29-OCT-1998; 98US-0106073.
PR 29-OCT-1998; 98US-0106074.
PR 29-OCT-1998; 98US-0106087.
PR 02-NOV-1998; 98US-0106587.
PR 02-NOV-1998; 98US-0106588.
PR 02-NOV-1998; 98US-0107034.
PR 02-NOV-1998; 98US-0107035.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J:
PI WPI: 2000-350688/30.
DR N-PSDB: A28712, A28713.
XX
XX Chlamydia antigens and the proteins they encode, useful for
PT vaccinating against Chlamydia infections that affect the respiratory
PT tract
XX
XX Claim 13; Fig 23; 226pp; English.
XX
XX The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
CC C. pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins
CC in samples (e.g. via enzyme linked immunosorbant assay (ELISA)).
XX
XX Sequence 931 AA:
SQ

```

xx Disclousure: Page 999-1000: 1755pp: English.
 ps Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
 cc Chlamydia trachomatis (see 201425). The polypeptides can be used as
 cc vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 cc can also be used to control growth of the microorganism. Chlamydia
 cc trachomatis is responsible for a large number of diseases, e.g. eye
 cc diseases such as conventional trachoma, nongonemic trachoma,
 cc paratrachoma, and inclusion conjunctivitis; genital diseases such as
 cc nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 cc perihepatitis, Bartholinitis; pneumonia; in breast feeding infants;
 cc and venereal lymphogranulomatosis. The polypeptides of the invention
 cc may be of use in treating these diseases.
 xx Sequence 989 AA:

Query Match 99.2%; Score 5047; DB 20; Length 989;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 958: Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKKAFFFLLNSGLAREVPSRFLMPNSVPDPKESLSNKISLTGDTNNTNCLYDN 60
 Db 25 mkkaiffllsisslsglarevpsrflmpnsvdpdkeslsnkisltdhnlnclydn 84
 QY 61 LRYTIAIIQKTPNEGAATITDYLSPDQEGIFYPAKNTPEEGCATGYASPSPTVEI 120
 Db 85 lrytiallqktpnagaavtictdyisftcdqeglyfaknlpsesgaglyaspsptvei 144
 QY 121 RDTIGPIYFENNTCCRPFTSSNPAAVKKIREGAIHQAOLYINIMHNDVYGFMKNFYVR 180
 Db 145 rdtigpylfemntccrpfstspnaavknkiregailhaqnllyinhndvgyfmknfysvr 204
 QY 181 GGATSTANTPVSNOSCFIFMDNTCIGTNTAGGAGAIYACTSNPSFSSNCDLFFINNAC 240
 Db 205 ggaistantlvsengscfifmdnclqtlagkgaalyagtsnfesncdlffinnac 264
 QY 241 CAGGATFSPICSLTGNRCNITVFYNNRCFKNETASSEADGAIKVTTRLDVNTNRGRIF 300
 Db 265 caggatfspicstlgnrnlvfyntcfknetasessoggaikvtrldvntgrif 324
 QY 301 FSDNITNNYGAIAPVYTLVDNCPYFINNIANNKGAIIYIDGTSNSKISADHAIIFN 360
 Db 325 fsdnltknysgaalyapvltvdnqpyffinnannkgaalyidgtsnksisadhaiifn 384
 QY 361 ENITVNTNNGTSTANPPRRNAITVASSSGELLGAGSSQNLITPDPIFVSNAGSVS 420
 Db 385 enitvntnngtstsanpprrnaitvaasssgellgagssqnlltfdplevsnagsvs 444
 QY 421 FNNKADOTGCVNFGATVNSADFHQRNLQKTPAPLTLNSGFLCIEHQAOLTVNRFTQTG 480
 Db 445 fnnkadotgsvnfsgatvnsadfhqnrnlqktktpapltlntsgflciehqaoltvnrftqtg 504
 QY 481 GYVSLGNGAVLSTCKNGAGNSASNAITLKHLNLSTLKSGAEIPLLWVEPTNNSNNY 540
 Db 505 gyvslngnavlscykngagsasnasitlkhlgnlslstksgaeipllwveptnnsny 564
 QY 541 TATTAATFSLSDVKLSITIDYGNSPYESTDLTTHALSSQPHLSTSEASDNOLKSDMDPSG 600
 Db 565 tattaatfslsdvklslidrygnspyestdlthalsqphlstsseasdnolrdsmdpsg 624
 QY 601 LNVPHYGMQGLMTWMAKTDPEPASATITDPQKANRFRITLTLTWLPAGYVSPKXHS 660
 Db 625 lnvphygwgglwvgaektdpepasatitdpkkanrfrtlltlwlpagvyppskhs 684
 QY 661 PLANTLMLGMLLATSLSKASALTPSDHPFWGITGGGLQWVYQDPRENIHPCFHMRSRG 720
 Db 685 pllantlmgmllateslksaeltpsdhpfwgltggglqwmvyqdprenihpfhmrsrg 744
 QY 721 YSAGMIAGOHFSLFQSOTYTKLNERVAKNNVSSKNSCOGEMLSLOGGFLTKLVNGL 780
 Db 745 ysa,miaagqthtstlksfslqtklneryaknnvssknscoqemlslslogfllaklvgl 804

QY 761 YSYGHNCHHEFTQGENLTSGCTFRSQTMGAVFEDLPKPGSTHILTAFLGALIGYS 840
 Db 805 ysygdnchhfytggenltsgctfrsqtmgavfddlpmpkpgsthilltaplgaiglys 864
 QY 841 SLSHTEGAVPRSRSTTPLINLVPIGVKGFPMNATQROPAWTEIAYQPVLYROEPG 900
 Db 865 slshltegavprststklplnvlpigvksfpmnatqropawteiaypvlyrgele 924
 QY 901 IATQILASKGIWFGSGSPSSRRHMSYKRSOQTPISWTLTFQVYGFSSSTFCVYLWGE 960
 Db 925 iatqilaskgiwfgsgspssrrhmsyksqtpiswlthfvygfssstfcvylwge 984
 QY 961 IALRP 965
 Db 985 ialrf 989

RESULT 2
 BI3634
 ID BI3634 standard; Protein; 964 AA.
 AC BI3634;
 DT 02-FEB-2001 (first entry)
 XX C. trachomatis pmpe gene protein.
 DE Chlamydia infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 XX Chlamydia trachomatis.
 OS
 PN W0200034483-V2.
 PD 15-JUN-2000.
 XX
 PF 08-DEC-1999; 99MO-US29012.
 XX
 PR 08-DEC-1998; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skelky YAM, Fling SP, Jen S, Stromberg EJ;
 XX WPI; 2000-431303/37.
 DR
 XX
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence
 PS Claim 2: Pages 184-186; 256pp; English.
 XX
 CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC infectious for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumoniae is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the
 CC present invention.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 19:22:01 ; Search time 22.61 Seconds

(without alignments)
2668.675 Million cell updates/sec

Title: US-09-677-752-4

Perfect score: 4533

Sequence: 1 MRPDHMFCCALCALISTYA.....CVLRGSGSHSYLDLGTTRF 878

Scoring table: BLOSUM62

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4509	99.5	878	2	B71460
2	3543.5	78.2	867	2	F81721
3	947.5	20.9	928	2	B72077
4	919.5	20.3	930	2	A81591
5	915.5	20.2	930	2	D72078
6	894.5	19.7	1276	2	C81591
7	894	19.7	1407	2	B72078
8	883	19.5	928	2	G81591
9	845	18.6	928	2	D72077
10	845	18.6	949	2	F81591
11	816.5	18.0	973	2	F72076
12	816.5	18.0	995	2	C81593
13	773	17.1	936	2	C72078
14	772	17.0	936	2	B81591
15	758	16.7	987	2	H81722
16	758	16.7	1013	2	G71460
17	740	16.3	922	2	B72131
18	739	16.3	922	2	F81539
19	708.5	15.6	841	2	E72130
20	575.5	12.7	946	2	D81594
21	574.5	12.7	946	2	C72075
22	566	12.5	947	2	D72067
23	557.5	12.3	978	2	B81593
24	556.5	12.3	978	2	G72076
25	540.5	11.9	938	2	H72074
26	531	11.7	934	2	H72075
27	531	11.7	952	2	D81593
28	510.5	11.3	1520	2	A81731
29	503.5	11.1	1609	2	H72013

30	493.5	10.9	986	2	B81675	polymorphic membra
31	487	10.7	1016	2	H71460	probable outer mem
32	484.5	10.7	975	2	F71518	hypothetical prote
33	482.5	10.6	976	2	F81722	polymorphic membra
34	475	10.5	964	2	E71460	probable outer mem
35	463	10.2	983	2	A81723	polymorphic membra
36	453	10.0	1531	2	H71468	probable outer mem
37	451	9.9	1034	2	F71460	probable outer mem
38	449.5	9.9	1025	2	G81722	polymorphic membra
39	445.5	9.8	1732	2	C81601	polymorphic membra
40	444.5	9.8	1723	2	E72067	polymorphic membra
41	413	9.1	1672	2	C81675	polymorphic membra
42	391	8.6	1751	2	G71518	hypothetical prote
43	341	7.5	1460	2	B81675	polymorphic membra
44	339.5	7.5	514	2	E72076	polymorphic membra
45	287.5	6.3	1770	2	A71517	hypothetical prote

ALIGNMENTS

RESULT 1
B71460
probable outer membrane protein I - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: B71460
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: B71460
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-878 <ARN>
A:Cross-references: G8:AE001361; G8:AE001273; NID:93329348; PIDN:AAC68472.1; PID:9332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: pmp1

Query Match 99.5%; Score 4509; DB 2; Length 878;
Best Local Similarity 99.5%; Pred. No. 5,7e-27;
Matches 874; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY	1	MRPDHMFCCALCALISTYAVLFGDDPLGETALLTKRPNHVVCTFEEDCTMESLFPALCA	60
DB	1	MRPDHMFCCALCALISTYAVLFGDDPLGETALLTKRPNHVVCTFEEDCTMESLFPALCA	60
QY	61	HASDDPLVVLGNSYCWPFVSKLHTDPKELFKKGGDLSIONFRFLSTDCSSKESP51	120
DB	61	HASDDPLVVLGNSYCWPFVSKLHTDPKELFKKGGDLSIONFRFLSTDCSSKESP51	120
QY	121	IQHKGQSLIRNNGSWFCRNHAECSGAIISADAFSLQHNYLFTAFEESSSKNGCAIOA	180
DB	121	IQHKGQSLIRNNGSWFCRNHAECSGAIISADAFSLQHNYLFTAFEESSSKNGCAIOA	180
QY	181	QTFSSLRNVSPIISFARNRDLNGAICCSNLICSGVNNLFTFGNSATNGAICCSIDLN	240
DB	181	QTFSSLRNVSPIISFARNRDLNGAICCSNLICSGVNNLFTFGNSATNGAICCSIDLN	240
QY	241	TSEKGSLSLACQOETLFFASNAKKEGAIYAKHMYLRYGAPSFINSKIGGATIAIOG	300
DB	241	TSEKGSLSLACQOETLFFASNAKKEGAIYAKHMYLRYGAPSFINSKIGGATIAIOG	300
QY	301	GSLSILAGGSVLPFONNSORTSDOGLVRNAIYLEKDALISLEARNGDLTFDPVIOESS	360
DB	301	GSLSILAGGSVLPFONNSORTSDOGLVRNAIYLEKDALISLEARNGDLTFDPVIOESS	360
QY	361	SKESPLSSLSQSVSPPTATASPLYIOTSNRVSIFSSERLSEEEKTDNLISQLOQT	420
DB	361	SKESPLSSLSQSVSPPTATASPLYIOTSNRVSIFSSERLSEEEKTDNLISQLOQT	420

Query Match 20.9% Score 947.5; DB 2: Length 928;
Best Local Similarity 31.6% Pred. No. 7.2e-52;
Matches 288; Conservative 137; Mismatches 336; Indels 149; Gaps 32;

```
85 TDPEKALFKKGGSLIONFAF--LSFTDCCSKSSPSIIHOKNG-QLSLRN----- 132
86 TNDGTTNTGTGVSITNMACSPALNASCRTKGTGSLFPGHGFQLQIYIDGANCFTT 111
133 -----NGSMSE-----CRNHAEGSGGAIADAFSLQHNLYLFAFEENSSKNGGAI 178
112 NTAAANLLSSGFSYLSLIOTFMATGTGAIKRTGACSIOSNV--SCFGQGNFNDNGAL 170
179 QAOFTSLRNVPISFARNRADLNGAICCSNLICSGN-VNPLEFTGNSATN----- 229
171 QGSSISLSLNPN-LTFAKNKATOKGGLYSTGGITTNNTLNSASFSENTAANNGAIYTE 229
230 -----GGAICISDLNLSFKGSLSLACQETLFAFNSAKK 265
230 ASSEISSNKAISTINNSVATSTAGATYSS--TSAPKPVLLSDGELNFTGNATITS 287
266 GGAIVAKHMYLRNGVPSFINNS-----AKIGGAIQSGSLILAGESSVLFQNNNS-- 318
288 GGAITYDNVLVSSGPTLFFKNNSAIDTAAPLGAIAIADSGSLSLAGDITFEGMTTV 347
319 -QRTSDGVLRNAIYL-EKDAISSLERNNGD-ILFEDPIVOESSKESPLSSLOASVT 375
348 KGAASSQTTNRNSINIGNTAKIYOLRASOGNTIYFDPIT-----TSITALLS 396
376 SPPTAPASPLVIQTSANRSVIFSESRLESEKKT-PDNLTSOLOOPTELKSGRLVLMKRAV 434
397 DALNLGPPDLAPPAQGTIVTFSGEKLSEAEALADNLKSTIOPPLLAGOGLSKSGVT 456
435 LSNPSSLQDPOALLIMEAGTSLKTSDDLKATLSIPLSLDTEKSVYIHAPNLSIQKIFL 494
457 LVAKSTISQSGSTLLMDAGTLETAAGTITNNLVANVDSLKETKATLAKKATQAS-QIVTL 515
495 SNS-----GDENFEYENVELLSKQNN--IPLLTSLKEQ-SHLHLPDGNLSS----- 537
516 SGSLSLVDPSPGNYEDV-----SMNPNQVESCULTTADDPANIHITD--LAADPLEKNPI 568
538 HFQYQDWTFSKMDSDSGHSLIA--NMTPKNVYVPHPEROSTIVANTLMNTYSMDQAQSM 595
569 HMYQGNMALSMQEDATATKSKAATLTWTGTGYNPNRKRGTLVANTLWGSFVDRSTIOQL 628
596 INTIANGAVLFGTWSAVSNLFYANDSSGKPIDNMHHRSLGVLFGISTSLDHSFCLA 655
629 VAKTVASQSETRGIMCEGISNFF--HKDSTRIKKGFRHISAGYVVGATTTLASNLITAA 686
656 AGOLLKSSDSFTSTETTSYIATVOAQ-LAT-----SLMK-----ISAQACYN 698
687 FCOLFCKDRDHFINKNKAAYAAASHLOHLATLSPSLLYRLPGESESEQVYLFQAQISYI 746
699 ESTHELKTYSKRSKRGFGSGMHSVAASGEVCASTPIYS-NGSGIFSSFSIFSKIOGSGT 757
747 YSNNTKMTYTOAPK-GESSWYNDGCALELASSIPHTALHHEGLFHAHYEPFIKEASYIH 805
758 QDGFEE-SSGEIRFSFASFRNISLPIGITFEKKSKQKTRFYVFLAYIIDLKRDVSGP 816
806 QDSFKENNTLVASFPGDGLINVSVPIGITFEHFRSERNRASYETTYIVADVYRKNDCT 865
817 VLLKNAVSDAPMANDSA-----YMRLLTNORALHRLQTLINVCVLRGQSHSY 868
866 TALLINNTSKRTIGTINLSRQAGIGRAGIIFYAFS-----PNLEVTSNLMEIRGSSRSY 918
869 SIUDGTYRF 878
919 NADLGKGFQF 928
```

RESULT 4
AB1591

polymorphic membrane protein G family CP0307 (imported) - Chlamydia pneumoniae (S
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: AB1591
R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzbe
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; M0ID:20150255
A:Accession: AB1591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-930 <REA>
A:Cross-references: GB:AE002193; GB:AE002161; MID:q7189234; PIDN:AAF38164.1; PID:q718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0307

Query Match 20.3% Score 919.5; DB 2: Length 930;
Best Local Similarity 31.0% Pred. No. 4.3e-50;
Matches 260; Conservative 150; Mismatches 343; Indels 87; Gaps 26;

```
98 LSIQNFRLSF-----TDCSSKSSPSIIHOKNGQLSRNNGSMSPCR--NNAEGSGA 149
119 LTFPGFENLSFLIAPCTTVASGKSTLS-----SAGALWLDHGTILFSQNSNEANNNGA 174
150 ISADAFSLQHNLYLFAFEENSSKNGGAI-QAOTFSLSRNVPISFARNRADLNGAICC 208
175 ITAKTLISGNTSSIFFTSSAKKLGAIYSSAAASISGNTGQLVFNNNGETGGAL-- 232
209 SNLICGNNV--NPLEFTGNSATN-----GGAICISDLNLSFKGSLSLACQETLFAFNS 261
233 -GEFASISITQNSLSLFFSGNTATDAACKGAIYC---EKTGEPTTLISGKNSLTFFENS 288
262 AKEGGAIVAKHMYLRNGVPSFINN-----SAKIGGAIQSGSLILAGESSVLFQNFON 316
289 SVTQGAICAHGIDLGAAGPTLFSNNRCNTAKGGAIAIADSGSLSLAGDITFEGL 348
317 NS-QRTSDGVLRNAIYLEKDAISSLERNNGD-ILFEDPIVOESSKESPLSSLOASV 374
349 NTLTSTAPSTRNAIYLGSSAKITNLRANQGGIYYDPIASMTTASDVL----- 400
375 TSPPATASPLVIQTSANRSVIFSESRLS--PEEKTPDNLTSOLOOPTELKSGRLVLMKRA 433
401 TINOPDSNPL---DYSGIVTFSGEKLSDAFAADNFTSILKOPALASGLTALGNV 456
434 VLSAPSLQDPOALLIMEAGTSLKTSDD-LKATLTSIPIHSLDTEKSVYIHA--N 486
457 ELDVNGFTQTEGSTLLMQPGTKLADTEALSTKLVVDLALBGNKSVSLETAGANKTIT 516
487 LSIQKIFLNSGDENFEYENVELLSKEQNNIPLTLTSKEQSHLHPDGNLSS-----HF 539
517 LTPSLIVQDSSG--NFESHTINQATPOLVFTFAATAADIIY-DALLTSPVQTPPEHY 573
574 GYQGHWEATVADISTAKSGMTWTWTTGYNPNPERASVYVDSLWASTDIRTLOOIWTSQ 633
540 GYQGDWTFSSKKSDSGHSLIANMTPKNYVPHPEROSTIVANTLMNTYSMDQAQSMINTI 599
600 AHGCAVLFQGWGSAVSNLFYANDSSGKPIDNMHHRSLGVLFGISTSLDHSFCLAQOL 659
634 ANSLYQORGMASTANFPH-KDKSGTN-QAFRKSIVGYVVGSAEDPSENIFSVAFQOL 691
660 LKSSDSFTSTETTSYIATVOAQ-----LATSLMK-----ISAQACYNES 700
692 FGKDKOLFIVENTSHNYLASLYLOHRAFLGGLPMPFSGSITDMKIDPLLNQLSYST 751
701 IHELKTYSRFSKRGFGSGMHSVAASGEVCASTPI--VNSGCLFSSFSIFSKLQFSGTOD 759
752 KNDMDITRYTSY-PEAQGSMNNSGALGLSLALYLTPKEAPFFQGYPPFLKFAVYSROQ 810
760 GFEESGEIRFSFASFRNISLPIGITFEKKSKQKTRFYVFLAYIIDLKRDVSSGTVL 819
```



```

0Y 155 PSLONNYLFTAEENS-KONGCAIQOFTSLNSVSP--ISPARNRDLNGAICCSNL 211
Db 179 FLLSGTSQFASFSRMOAFTCKQGVVATGITYITL ENSFGIYFSFONLAKGSGALYSTD- 237
0Y 212 ICSGNVN-PLUFTGNSA-----TNGCAICCSIDNTSEKGLSLACNOETLEFASNAKEK 265
Db 238 NCSITDNFQVIFPDGNSAMEAQAOGGAICCC---TTDCK-LVTILGNKLSFTNNLTALTY 292
0Y 266 GGAIIAKHAWLRYNGVSPFINNSAKT-----GCAIIAIOGSGSLSLIACEGVILQONNQ 319
Db 293 GGAISGLKLSIASGGPTLTFUSNISGSSAGOGGAINASAGELALSAATSGDITF-NNNQ 351
0Y 320 RPSDGLVNAIYLEKDALISLEARNG-DILFDPPIYOESSKESPLPSLSQASVTSPT 378
Db 352 VTNGSTSTRNAIIVITDAKATSIINAATGOSIYFYDPIITNPGTAATDPLNLNLADANS-- 409
0Y 379 PATASPLVIQTSANRSVIFSSERLSEEEKT-PDNLTSOLOOPIELKSGRLVAKRAVLSA 437
Db 410 -----EIEYGAIVFSGEKLSPEKAIANAVTSTIRQPAVLARGVLARDGTVTF 460
0Y 438 PSLISODPQALLINEAGSLKT-SSDLKATLSTPLHSIDTEKSVIHA-----PNLSIOKI 492
Db 461 KDLTOSPGSIILDDGGTTLAKENLNLGLAVNLSSLDGTNKAALKTEADKNISLSGT 520
0Y 493 PLNSGSDENYENVVELLSKEONNIPILTLTSKEOSHLLHPDGNLS-----SHEGYOGD 544
Db 521 IALIDTEGSEYENHNL--KASATYPLELTTAGANGITIGALSTLTLOEPETHYGOGN 578
0Y 545 WTFSMKOSDEGSHLIANWTPKNIYHPPEROSTLVANTIMNTYSMDQAVOSMINTIHGCA 604
Db 579 WOLSMANATSSKIGSIMTRGTYIPSEBKSNLPLNSLWGNIDIRSIQWLIETKSSGP 638
0Y 605 YLEFGTGSANSNIFYANDSSKPIDNMHRSIGYFPGISTHLDHSPFLACOLLGKSS 664
Db 639 FERELWLSGANFPRDSMPTR--HGFNHSIGGVALGITATTAPBDQLTFACQLLARDR 696
0Y 665 DSFTTSTETTSIATVOALATSLMKIS-----AOACYN 698
Db 697 NHITCKNHGDTYGASLVEHTEGLFDIANPLMKATRAPVWLSLSQIILPSFDAKFSYL 756
0Y 699 ESTIHELKTKY--NFSKSEGFSGSHSVANSGEYCASIPIYNSNGLSFSEIFKLGFGSG 756
Db 757 HTDNHMKTYYTDNSIIR--GSMRNDACADGASLPEVISVYLLKEVEPEFYKQIYYA 813
0Y 757 TODGFEESSGEIRFSFASFPNISLPIGITFEKKKSQKTRTYVYFLCAYIIDLKRDVESP 816
Db 814 HOODFYERHAEGRAFNSSELINVEIPIGVTPERDRSKSEKGTDLTIMYILDAIRBNPKCO 873
0Y 817 VLLKNVNSMDAPRANLDSRAVMEFLTNORALH-RLQTLNNSCVLURGOSHYSLDLGT 875
Db 874 TSLIASDANMAYCTNLAKOCFSVRAANHGVQNPMEIIFGOFAEVHSSSRNNTNLGSK 933
0Y 876 YRF 878
Db 934 FCF 936

```

RESULT 14
B81591
polymorphic membrane protein G family GP0308 [imported] - Chlamydophila pneumoniae (streptococcus)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81591
R:Reed, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
. C.; Dodson, R.; Gunn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: B81591
A>Status: preliminary
A:Molecule type: DNA
.:Residues: 1-936 <REN>

A:Cross-references: GB:AE002193, GB:AE002161, NID:g718923, PID:AAFC38165.1, PID:g718
A:Experimental source: Strain AR39, HL cells
C:Genetics:
A:Gene: CP0308

Query Match	17.0%;	Score 772;	DB 2;	Length 936;
Best Local Similarity	28.0%;	Pred. No. 9e-41;		
Matches 236;	Conservative 133;	Mismatches 378;	Indels 86;	Gaps 22;

```

QY 97 DLISIONFELISFTQCCSSHESSPS--IIHONKCOLSLRNNNSMFCNHHAGSGCAIADSA 154
Db 119 NLEFNDFEURLISICSPSLSPITQCALKSYSVNLSTGNSOIIIFTFONFSSDNGGVYLTNK 178
QY 155 FSLONHYLFTFAFEENSS-KNGNGAIOAOTFELSFRNVP--ISFARRADJLNGAICCSNL 211
Db 179 FLIISCTOSFASFSRRQATGTCOGGVYATGTTTENSPLGVFSQNLANGSGALYSTD- 237
QY 212 ICSGNVN-PLFEFTGNSA-----TNGAICCTISDLNTESEKSLSLACNOETLEFASNAK 265
Db 238 NCSITDNFOVIFDGSASAEAAOAGGCAIC-----TTTDDK-TVMTGNKNISFTNNTALTY 292
QY 266 GCATYAKKHVILRYNPPVSFTNNSAKT-----GGAIIOSGCSLSIAGCGSVLPNNNSO 319
Db 293 GGAISGLGVYSISAGCPPTLFQONISGSSAGCGGGAIIAAGELALASVTSGDITF-NNNO 351
QY 320 PRSDGCLVRNATYLEKDAIISLEARNG-DTIFPDPIVOESSKESPEPSLSOASVTSPT 378
Db 352 VINGSTSTRNAINIIDTKAVYSTRATGQOISITFFDPTINTGTASTDYTLNLNLADANS-- 409
QY 379 PATASPLVYIOTSANRVSIFSESRLESEERT-PDNLTSOLOOPIELKSGRLVLDRAVLSA 437
Db 410 -----ELEGGCAIYFSGEKLSPREKALIAVANVSPRIQPAVLARGDLVLDGYTVAF 460
QY 438 PSLISODPQALLIMEGTSUKT-SSDKIATISITPLSHLIDTEKSVTHA----PNLSIOKI 492
Db 461 KDLTOSPGSRILMDGTTLSLAKKEANTSLNGLAIVNLSLSDGTNKAALTEADNRNLSISGT 520
QY 493 FLTNSGDEAFYFNVELLSKEDONNIPULTLSKEQSHLHPDGNIS-----SHFGYOG 544
Db 521 IALIDTEGSFYENHNL--KSASTYPLPLITTAGANGITTYIGALISTLLQEPETHYGTQGN 578
QY 545 WTESMKDSEGHSLIANNTPKNYVHPEROSTLIYAANTLMTNYSDMAQVQSMNTIAHGA 604
Db 579 WOLSAANAATSSKIGSLNMTGRICYISPERKENLPLNLSMGNFIDIRSLNOLIEKSSGEP 638
QY 605 YLEGTWGSASVNIIFYAHDSCKP-IDNMWHRSILGYLFGISNLDHSHFCLACQULGKSS 664
Db 639 FERELTSLCIANFFRDSMPTR--HGFRHISGGVALGITATTPAEDOLTFACOLFARDR 696
QY 665 DSFTSTETTSYIAVQOALSLMKIS-----AQCYN 698
Db 697 NHTGKNHGDITGASLYFPHTEGLRDLIANFLMGKATRAWPLSEISOIPLSPDAPKSYL 756
QY 699 ESITHELKTRY--RSEKKEGFGMHVAVASGEVCAISPIVNSGSLJFSSSIFSKLQFSG 756
Db 757 HTDNHAKTYTNDNSLIK---GSMRNDACADIGASLAPRIVISVPLLKEVEFEFVVOYIYA 813
QY 757 TODGFEESGELIRFSBASSFRNISLIPCIYTEKKSQKTRTYVYLGAYIODLKRDVESGP 816
Db 814 HODDFERYAEGRAFNKSELLINVEIPIGTEFERDSKSEKGYDLTLMAYIIDAYERNPKQ 873
QY 817 VLLKAAVSWDAPMAUDSRAMPFLUTQORAH-RLOTLNLAVSCVLNGQSHSYSLDITGT 875
Db 874 TSLIADDAWMAAGTNLAROGFSVRAAHFPOVNPHELFQGFALVEVSSSKNNTNIGSK 933
QY 876 YRF 878
Db 934 FCF 936

```

RESULT
H81722

Db 506 -----NGNSTLYONTTIEOGRIVLEKAKLVNSLSUTGGS-LYMEAGSTIDFV 553
Qy 457 -----KTSDDLKIALTSLPLHSI-----DREKSVTI 482
Db 554 TTPROPPOPPANOLITLISNLHLSSLLANNAVTPPTNPAPADSHPAIIGSTTACSVTI 613
Qy 483 HAPNLSTOKIFLSNGDENFENVELLSKEON-----NIPULTLSKEQSHLPLPGNLSH 538
Db 614 SGP-----IFEDIL-DOTAVDRYDMLGSNOKIDVLKIQLOLQPSANASDITLGENMPK 666
Qy 539 FCYOGDMTFW--KQSDG--HSLIANMTPKNVVPHPEPOSTLVANTLNNTYSDMOAVOSM 595
Db 667 YGQGSWMLANDPNTANNQPTLKATWTKTGTNPPCPERAVASLVPMSLGSILIDISANSA 726
Qy 596 INTIAHGAVLFTGWSAVSNLFVAHDSGKPIDMHHRSILGYLEFSTHSLDDHSPCLA 655
Db 727 IQASVDGSHYCGGLWVSVSNFFY-HDRDALC-QGYRYISGYSIGANSY-FGSSMFGLA 783
Qy 656 AGOLIGKSSDSFTISTE-----TTSYIATVOQLATSLM-KISAQACYNESHELKTKR 709
Db 784 FTEVFGSKRDYVVCRRSNHACIGSYLLSTKQALCGSYLFGDAFIRASYGFGQNMKTST- 842
Qy 710 SFESKGFSGWHSVAVSGEVCASIPIVNSGGLF-SFSIFSKLOGFSGTODGEESGEBI 768
Db 843 TFAESDVRMDNCLVGEIGVGLPIVITPSKLYLNEIRFVQAEFSYADHESFTGEGDA 902
Qy 769 SFSASFSRNISLPIGITEFKKSOKTRTYVEFLGAYIQDLKRDVESGPVLLKNAVSWDA 828
Db 903 RAFRSGHLMNLSPYGVKRCDSSTHPNKYSFGAVICAYRTIISOTQYTLTSHQETWTT 962
Qy 829 PMANDSRAYMRLTNORAL-HRLQTLNVCVLGQSHSYSLDGTYYRF 878
Db 963 DAFHLARHGVIIVGSMYASLTNSIEVYGHGRYEYRDTSRGYGLSAGSKVRF 1013

RESULT 17
B72131
polymorphic outer membrane protein 9 family - Chlamydothilla pneumoniae (strain CWL029)
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: B72131
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: B72131
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <ARN>
A:Cross-references: GB:AE001585; GB:AE001363; NID:94376235; PIDN:AND18163.1; PID:9437626
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: pmp_1

Query Match 16.3%; Score 740; DB 2; Length 922;
Best Local Similarity 27.2%; Pred. No. 9.3e-39;
Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

Qy 103 FRFLSFTDCSKESPSTIIOKNQOLSLRNNGMSFPCRNIAEESGGATISADAFSLQHNLY 162
Db 123 FSTLSFTIOSPDIGKEQCLYSKNA-LMLNNYVVRFEONOSKTRGKAIISGANVTIYGNVD 181
Qy 163 FTAFENSSKNGNGAIOAOTFSLSRNVSPISFARNRA-DLNGCAIIC-SNLIJSGNVNPL 220
Db 182 SVSTQYNAALFGCAIHSSGLOIATVANOALIRFQNTAKNGSGALYSDGDIQDNMYVL 241
Qy 221 FTGNSAT-----NGCAICGISDLNTSEKSL-SLACNOETLFPASNAKKEKGAIYAKHNV 275
Db 242 FREHEALTTAIGKGAACCLPTSGSFPVPDVTFSDNKQLVFERNHNSIMGCGAIYARKLS 301
Qy 276 LRNGPVSFIN-----NSAIGCAIAIOSGGSILAGEGVLFQNNNSQRTSDQGLVRNA 330

Db 302 ISSGPTLFINNISYANSONLGAIADTGCSELSAEGTITTECGN--RTSLPFL--NG 357
Qy 331 IYLEDALISSLEARNG-DILFPPIVQESSKESRPLSSIQASTPTPTAPASPLVIQT 389
Db 358 IHLLONAFLKIQARNNGSIEFYDITSEAD-----GSTQLNNGDPK-----NK 402
Qy 390 SANRSVFSERLSEETPDNLTSQLQPIELKSGRLVLDRAVLSPASLQDPAOLI 449
Db 403 EYTGITLFSGEKSLAND--PRDFKSTIPQNNVLSAGLYIVINAGAVTYSKTFQSGSLV 460
Qy 450 MEAGTSLKTS-SDKLATLSPILSHSLDTEKSVTHAPNLSTOKIFLSNGD-----ENFY 503
Db 461 LDGKTLASKEDIAITGAIDIDLSSTSTAAVYKANTANKQSVTSDIELISPTGNAY 520
Qy 504 ENVELLSKEQNNIPLTLTSK-BQSHLHLPDGN-----LSHFQYOGDMTFWMSKDSDE--GHS 557
Db 521 EDLRM--RNSOTFPLLSLEPGAGSVTYAGDFLPVSPHYGQGMWKIAMGTGKKBGF 578
Qy 558 LIANMTPKNVVPHPEPOSTLVANTLNNTYSDMOAVOSMINTIAHGAVLFTGWSAVSNL 617
Db 579 F---WDKINRPREKEGCLVPLNLLMGNAVDRSLMQVETHASSLQDRGLMDIGMVF 635
Qy 618 FYAHDSGKPIDMHHRSILGYLEFSTHSLDDHSPCLAQOLIGKSSDSFTISTETSYI 677
Db 636 F--HYASASEDNIRYRHNSGGLVSVNNEITPKHYTSMASFQLSRDKDYAVSNNEYRML 693
Qy 678 ATVOQLATSLMKISAQA-----CYNESHELKTKYRS 710
Db 694 GSILYQYTLSTLQINFRYASRNPNVNGILSRPFLQNLIFHLPCAYGATMDMTIDIAN 753
Qy 711 FSKEGFGSWHSVAVSGEVCASIPIVNSGGLF-SFSIFSKLOGFSGTODGEESGEBI 769
Db 754 FPMVK-NSKRNKMCALTECGSMPLVLFENGRLFGQALIPPMKIQLYAVAGDPRKETAADR 812
Qy 770 SFESASFSRNISLPIGITEFKKSOKTRTYVEFLGAYIQDLKRDVESGPVLLKNAVSWDA 829
Db 813 RFSNGSLTISVPLGIRFEKFLALSDQVLDVDFSEFYPIDIFRKDPSCALVLVSGDSMLVP 872
Qy 830 MANLDSRAYMRLTNORALHRLQTLN--NVSVLRLGQSHSYSLDGTYYRF 878
Db 873 AAHVSHARVAGTGRVYHFNDDYELLRCGSIETC--RPHARNYNINCGSKFRF 922

RESULT 18
F81539
polymorphic membrane protein G family CP0770 (imported) - Chlamydothilla pneumoniae (s
C:Species: Chlamydothilla pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: F81539
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hicke
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: F81539
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-922 <REA>
A:Cross-references: GB:AE002237; GB:AE002161; NID:g7189684; PIDN:AAF38570.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0770

Query Match 16.3%; Score 739; DB 2; Length 922;
Best Local Similarity 27.2%; Pred. No. 1.1e-38;
Matches 226; Conservative 156; Mismatches 362; Indels 88; Gaps 25;

Qy 103 FRFLSFTDCSKESPSTIIOKNQOLSLRNNGMSFPCRNIAEESGGATISADAFSLQHNLY 162
Db 123 FSTLSFTIOSPDIGKEQCLYSKNA-LMLNNYVVRFEONOSKTRGKAIISGANVTIYGNVD 181
Qy 163 FTAFENSSKNGNGAIOAOTFSLSRNVSPISFARNRA-DLNGCAIIC-SNLIJSGNVNPL 220

RESULT 20
DB1594
polymorphic membrane protein E/F family CP0283 (imported) - Chlamydia pneumoniae (str
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: DB1594
R:Read, T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255
A:Accession: DB1594
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-946 <RNA>
A:Cross-references: GB:AE002189; GB:AE002161; NID:97189205; PIDN:AAF38141.1; PID:9718920
C:Genetics:
A:Experimental source: strain AR39, HL cells
A:Gene: CP0283

Query Match 12.7%; Score 575.5; DB 2: Length 946;
Best Local Similarity 23.4%; Pred. No. 2.4e-28;
Matches 231; Conservative 151; Mismatches 427; Indels 179; Gaps 33;

OY 13 AAILSTAVLEFGDPLGETALLTKNPNHYVCTFEEDCTMESLFP-----ALCAHASQ 64
DB 16 ALILGKTTILLNATPLSD--YFDNQANQLTTLPLIDTLTNTMPSHRATLFGVRDDTNO 73
OY 65 DDPLVYLGNSYCF-----VSKLHITDPKEAL-----FKKGDLSIQ-NFRF 105
DB 74 DIVLHQNSTIESMFENFSODGALSCSLAITNTKNOILFLNSFAIKRAGAMVNGNFDL 133
OY 106 -----LSFTDCSSKE---SSPSIIHOKNGQLSRNNGSMFCRNHAEAGSGAI 150
DB 134 SENHGSIIIFSGNLSPFNAFADCTCGAVALCSKNVTIS-KNGCTAYFINKKAKSSGAI 192
OY 151 SADAFSLQHNLYLETAFEENSSKGN-GGAIOAQTFSLSRNVSPISFARNRADLNGAICCS 209
DB 193 QAALIIKIDNTGCPCLFFNNAAGTAGALFANACRIENNNSQPIYFLFNNSGLGAIIRVHQ 252
OY 210 NLICSGNVNPLPFTGNSA-----TNGAICCIDLNTSEKGSISLACNOETLFSAN 260
DB 253 ECLITKNTGSIYFENNPFAMEADISANHSSGAIYICIS--CSIKDNPGLA-----FDNN 304
OY 261 SAKKGAATYAKHMLVRYNGPVSFINNNAKIGAIQSGSLSILAGEGVLFONNSOR 320
DB 305 TAARDGALCTOSLTIDSGPVYFTNNOGTWGAIMLRDGACTLPADGDIIFYNN-RH 363
OY 321 TSDGVLNMAIYLEKDALISLEARNGDILFDPPI-----VOESSKESPLPSLQASVT 375
DB 364 FKOTFSNHNVSVCNTRNVSLTVGASOGHSATFYDPILORYTIONSIOKFNPNPEHL----- 418
OY 376 SPPPATASPLVIOTISANRSVTFSSERISEBEKTPDNTSLOQPIELKSRVLVKDRAVL 435
DB 419 -----GTLFSSAVIPIDTSTRDPIISHFRHHIIGLYNTALIEDRAEW 461
OY 436 SAPSLSDOPQALLI-----MEAGSLKTSDDLKATLSTIPLHS----- 474
DB 462 KYVKPDPQGGTLRLGSHRAVFTTIDEESSSSVGSVININMLAINPLSLGNRAVAKIMIR 521
OY 475 -----DTEKSVTIHAAPNLSTOKIFLNSGDENF--YENVLLSKRONNIPIL-TL 521
DB 522 PTGSSAPYSDDNPIINLSP-----LSLDDENLDPIYTRAD-LAOPIAEVPFLYLL 572
OY 522 SKROSHLA-----LPDG-NLSHFQYOGDWTFPSKMD-----SPG-----HSLIANWT 563
DB 573 DVAKAKHINTNFYPEGINTTOHYGVGWSPYMETITTSSTJSESDYVNTLHROLYQWT 632
OY 564 PKNVVPEROSTLVANTLMTYSDM-----QAVOSMINTIAGGAYLFTGWSAVSNL 617
DB 633 PTGKVPKPNKGDIALSAFWSQFNNLFAITLRYOTQOQIAPTASGEA-----TR 681

OY 618 FYAHDSCGKPIIDMWHHRSLGYLFGISTHSLDHSFLACQILGKSSDSPTSTETTSYI 677
DB 662 LFVHONSNNDKAKGFHMEATGVSIGTTSNTASHSGVNFSQLFSLYESH-SDNSVAHHT 740
OY 678 ATVOALATSLMK-----ISAQACYNESIHELTKYRFSRKEGFGWSHVAVGEVCASIP 733
DB 741 TTVALQINNPMLQERFSTASLAYSYSNNHIIKASGYSKIQTEGCYSTTLGALSCSL 800
OY 734 IYSNNSGFLSSSTISKLDGFGCTODGPEESGCELRFSFA-SFFNISLPIGTFEKKSO 792
DB 801 LQWRSRPL-HETPEIQALVRSNOTAFQESGDKARKFSVHKPLVNLTVPLGDSQASWK 858
OY 793 -KTRTYVFLGAYIQDLKRDVESGPVLLKNAVSDAPMANLDSRAYMRLTNGRALH-R 850
DB 859 FLPTPYMTELAYQVLYOONEPVNVSLESSSSMLSTTLARNAIAFKGNQIFIPK 918
OY 851 LQTLNVCVLRGQSHSYSLDGTYYRF 878
DB 919 LSVPLDYQGSVSSSTTHYLHAGTTFK 946

RESULT 21
C72075
polymorphic outer membrane protein e/f family - Chlamydia pneumoniae (strain CWL0
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72075
R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206506
A:Accession: C72075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-946 <ARN>
A:Cross-references: GB:AE001631; GB:AE001363; NID:94376750; PIDN:AA018610.1; PID:9437
C:Genetics:
A:Experimental source: strain CWL029
A:Gene: pmp_18

Query Match 12.7%; Score 574.5; DB 2: Length 946;
Best Local Similarity 23.4%; Pred. No. 2.8e-28;
Matches 231; Conservative 151; Mismatches 427; Indels 179; Gaps 33;

OY 13 AAILSTAVLEFGDPLGETALLTKNPNHYVCTFEEDCTMESLFP-----ALCAHASQ 64
DB 16 ALILGKTTILLNATPLSD--YFDNQANQLTTLPLIDTLTNTMPSHRATLFGVRDDTNO 73
OY 65 DDPLVYLGNSYCF-----VSKLHITDPKEAL-----FKKGDLSIQ-NFRF 105
DB 74 DIVLHQNSTIESMFENFSODGALSCSLAITNTKNOILFLNSFAIKRAGAMVNGNFDL 133
OY 106 -----LSFTDCSSKE---SSPSIIHOKNGQLSRNNGSMFCRNHAEAGSGAI 150
DB 134 SENHGSIIIFSGNLSPFNAFADCTCGAVALCSKNVTIS-KNGCTAYFINKKAKSSGAI 192
OY 151 SADAFSLQHNLYLETAFEENSSKGN-GGAIOAQTFSLSRNVSPISFARNRADLNGAICCS 209
DB 193 QAALIIKIDNTGCPCLFFNNAAGTAGALFANACRIENNNSQPIYFLFNNSGLGAIIRVHQ 252
OY 210 NLICSGNVNPLPFTGNSA-----TNGAICCIDLNTSEKGSISLACNOETLFSAN 260
DB 253 ECLITKNTGSIYFENNPFAMEADISANHSSGAIYICIS--CSIKDNPGLA-----FDNN 304
OY 261 SAKKGAATYAKHMLVRYNGPVSFINNNAKIGAIQSGSLSILAGEGVLFONNSOR 320
DB 305 TAARDGALCTOSLTIDSGPVYFTNNOGTWGAIMLRDGACTLPADGDIIFYNN-RH 363
OY 321 TSDGVLNMAIYLEKDALISLEARNGDILFDPPI-----VOESSKESPLPSLQASVT 375
DB 364 FKOTFSNHNVSVCNTRNVSLTVGASOGHSATFYDPILORYTIONSIOKFNPNPEHL----- 418

Mon May 7 09:06:41 2001

Search completed: May 6, 2001, 19:26:18
Job time: 257 sec

us-09-677-752-4.rpr

OY 563 TPKNVPHPEROST-----LVANTLMNTYSMDQA 591
DB 677 TEIKYISSNSKGLTQYRSSAGVNFNGVNGNMSFNLKCGAKVNFKLKPRENNMT-SKPLP 735
OY 592 VOSMINTIAHGAYLFGTWGSAVSNLFYAHDSGKPID-NMHRSLGYLFGISTHSLDDH 650
DB 736 IRLPLNITATGG-----GSVFEDIVANHSGRGAELKMSLEINISGANFTLNSHYRGDD 788
OY 651 SFCLAGOLLKSSDSFSTSTETTSYATVOALATSLMKISAQACYNESHELKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLQTKDDFYD 816
OY 710 SFSKEGFGSMHSAVSG 726
DB 817 GYARNAINSTYNISILG 833

RESULT 3
US-08-530-198-4
Sequence 4, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERNSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-4

Query Match 3.3%; Score 151.5; DB 2: Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

OY 73 NSCYWVSKLHTDPRKALFKKSG-----DISTONRFSLFTDCSSKSS-PSLIHQ-- 123
DB 143 NGOVFLINPNCITIGDAIINTNGFTASTLSDISNENKARNFTEQTKDALAEIVNHGL 202
OY 124 -----KNGQSL-----RNNGSMSCFRNHAEGSGAIS-----ADAFSLQHNLYLFTA 165
DB 203 IYVGKGSVNLIGGKVKNEGVISV-----NGGSISLLAGQKITTSIDIIINPTITYSIAA 255

OY 166 FEENS-----SKG-----NGCAIQAOFFSLSRNVPISFARNADLNGCAIC 207
DB 256 PENEAVNLGDIKFAKGNINVRATIRNOGKLSADSVSKDGNIVLSAKEGEALIGVYS 315
OY 208 CSNLLSCGNVNPFFTCNSAT-NGCAICISDLNTESEKGSLSLACNOETLFRASNAK-K 265
DB 316 AONQAKG--GKLMITGCKVTLTKGAVI---DLSEKEG-----ETYLQGDGRCGK 362
OY 266 GCAITAKHMYLRYNCPVPSFINNSAKIGAIQSGSLSILAGESSVLPONNSQNTSOG 325
DB 363 NGIOLAKKTSLEK--STINVSKEKEGFAIVWG--DIALIDGINAQSGDIAKTGG 416
OY 326 LVRNA-----YLEKDALISLEARNCDILFFDPTVOESSKESPSLQASVTSPTPATA 382
DB 417 FVETSGHDLFIDKNAIVDAKER-----LDFEDVY-----STNAEDPL----- 453
OY 383 SPLVIQTSANRSVIFSSERLSEEEKTPDNI.TSLOOPIELK-----SGRL----- 427
DB 454 -----FNNMG INDEPFTGGEASDPKKNSELKTTLNTTISNLYKANMTWN 499
OY 428 VUKRAVLSAPLSODPQALLIM-----EAGSLKSSDLKLTLSIPHS---LDTEKSY 480
DB 500 ITASRLITVNSSINIGSNHLLHSKQKQGVQIDGDTTSKGNLTLYSGGWVPHKNI 559
OY 481 TIHAPLSTOKIFLSNGDENFYE--NVELLSK-----EONNIPV-----L 519
DB 560 TLDQGFLLNTAASVAEFGNNKARDAANKIYAOGVITTTGEGKDFRANVNLNGTKGL 619
OY 520 TLSKEQSHLPLDGNLSHFYOGDWTFES-----WKDSEGHSLI-----ANW 562
DB 620 NISSVNNL---THNLSGFINISGNTITNOTTKRNTSYQTSHDSHMVNSALNLETGANF 676
OY 563 TPKNVPHPEROST-----LVANTLMNTYSMDQA 591
DB 677 TEIKYISSNSKGLTQYRSSAGVNFNGVNGNMSFNLKCGAKVNFKLKPRENNMT-SKPLP 735
OY 592 VOSMINTIAHGAYLFGTWGSAVSNLFYAHDSGKPID-NMHRSLGYLFGISTHSLDDH 650
DB 736 IRLPLNITATGG-----GSVFEDIVANHSGRGAELKMSLEINISGANFTLNSHYRGDD 788
OY 651 SFCLAGOLLKSSDSFSTSTETTSYATVOALATSLMKISAQACYNESHELKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLQTKDDFYD 816
OY 710 SFSKEGFGSMHSAVSG 726
DB 817 GYARNAINSTYNISILG 833

RESULT 4
US-08-469-880-4
Sequence 4, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: OF NO. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:


```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO.: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1338 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-9

```

```

Db      881 AAGVTTTKEETTINATTGSGVEYTAQCKTICKNITGTSQNTVTATEKLVTTENAVINATSGT   940
QY      590 QAVQSMNTATAGGAILFTGWG----SAYSNLEFYAHDSGKRIIDMHHRLSLGLFGLSTH   645
Db      941 VNISTKTGDIKCG---IESYSGVNVITAGCNLUKKNITGOLV-----TV 982
QY      646 SLDDHSPFLACAGOLLCCKSSDSPTSTETTSYIATVQAOL-ATSLMKISAQACYNESIHIL   704
Db      983 TADGALTTTAGSTISATITGNANITTKTGDINKCKVESSGSVTLVATGATLAV----- 1035
QY      705 KTYRPSKKEGFCGSMHVAV---SGCEVASIPVNSGSLFSPSIFSKLQ-PEGTODG 760
Db      1036 -----GNISGNVITTADSGKLTSPGSGTNGTNSVTTSSOSGDIKCTISGNTVN 1085
QY      761 FEESGGEI 768
Db      1086 VTASTGDL 1093

```

```

1      RESULT 9
2      US-08-038-682-2
3      : Sequence 2, Application US/08038682
4      : Patent No. 5549897
5      : GENERAL INFORMATION:
6      : APPLICANT: BARENKAMP, STEPHEN J
7      : APPLICANT: ST. GEME II, JOSEPH W
8      : TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
9      : TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
10     : NUMBER OF SEQUENCES: 8
11     : CORRESPONDENCE ADDRESS:
12     : ADDRESSEE: Shoemaker and Matlare, Ltd
13     : STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
14     : STREET: Bldg. 1
15     : CITY: Arlington
16     : STATE: Virginia
17     : COUNTRY: U.S.A.
18     : ZIP: 22202-0286
19     : COMPUTER READABLE FORM:
20     : MEDIUM TYPE: floppy disk
21     : COMPUTER: IBM PC compatible
22     : OPERATING SYSTEM: PC-DOS/MS-DOS
23     : SOFTWARE: PatentIn Release #1.0, Version #1.25
24     : CURRENT APPLICATION DATA:
25     : APPLICATION NUMBER: US/08/038.682
26     : FILING DATE: 16-MAR-1993
27     : CLASSIFICATION: 424
28     : ATTORNEY/AGENT INFORMATION:
29     : NAME: BERKSTRESSER, JERRY W
30     : REGISTRATION NUMBER: 22,651
31     : REFERENCE/DOCKET NUMBER: 1038-293
32     : TELECOMMUNICATION INFORMATION:
33     : TELEPHONE: (703) 415-0810
34     : TELEFAX: (703) 415-0813
35     : INFORMATION FOR SEQ ID NO: 2:
36     : SEQUENCE CHARACTERISTICS:
37     : LENGTH: 1536 amino acids
38     : TYPE: amino acid
39     : STRANDEDNESS: single
40     : TOPOLOGY: linear
41     : MOLECULE TYPE: protein
42     : US-08-038-682-2

```

Db 554 WDVHKNISLGAQGNINITAKODIAFEKGSNOVITIGOGTITSGNOKGFFRNVSJLNGTGS 613
QY 665 DSFIETSTETSYIATVQAOIATSL--MKISAQACYNESIHLELTKYRSFSKGFSGSWH- 720
Db 614 GLOFTTKRNTKRYATNKKEGTINISGVNISMVLPKNEGYD-KFKGRY-----WNL 665
QY 721 --SVAVSGEVCASIPVNSGCLFSSFSIESKLOGFSGTODGFEBSGCEITSFASSTR 777
Db 666 TSLNVESEGEF--NLITDSRGSAGLITLOPYNLNGISFNKD---TTFVVERNARVNF- 718
QY 778 NISLPIGI 785
Db 719 DIKAPIGI 726

RESULT 11
US-08-530-198-2
Sequence 2, Application US/08530198
Patent No. 5869065
GENERAL INFORMATION:
APPLICANT: BARENKAMP, STEPHEN J
APPLICANT: ST. GEME III, JOSEPH W
TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,198
FILING DATE: 13-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: JWB-1186
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-530-198-2

Query Match 3.08; Score 134.5; DB 2; Length 1536;
Best Local Similarity 19.4%; Prid. No. 0.0058;
Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

QY 120 IIMOKNGOQLSRNNGSMFCRNHAECSGAIISADAFSLHNYLFTAFEEHSSKNGCAIQ 179
Db 76 VVH---GTATMOYDGNKNTITRNSVD---AIIIMKOFINIDONENKQFOEN---NNSAVF 125
QY 180 AGRFSLSRNYSPIFARNRADLNGAICCSNLCISGVNPLFFTGNSATNGAICISDL 239
Db 126 NRTYS-----NOISOLKGLIDSNQVFL-----INP-----NGITIGKDALI 162
QY 240 NTEKGSLSLACNOETLFSAN---SAKEKGALYAKH---MVLRYNGPVSEFINNSAKIG 293

Db 163 NTNGTASTLSDISNENIKRNFTEGOTKXALAEIVNHGLIVGKDGSVNLIGKVKNEG 222
QY 294 AIAIOGSGGLSLIAGEBSVLFONNNSORTSDQGLVNRNAIL-----EKDAI-LSSLFARG 347
Db 223 VTSV--NGGSIISLAG-----OKITISDIINPTIYISAABENEAIVNLGDIFAKGG 271
QY 348 DILFDPRIYQESSKESPLSSLOASVTSPTPATSPULIOTNANRSVYFSSERLSEER 407
Db 272 NI-----NVRATIRNOGKISADSV 291
QY 408 TPDNLTSLOQPIELKSGRLVLDKDR-----AVLSAPLSODPOLIMEAGTSLKTS 460
Db 292 SKD-----KSGNIVLSAKEGEALIGGVISAQNOQAKGGLMTGDKVILKTA 339
QY 461 DLKLA-----TLSTPL-HSLDEKSVTHAD----- 485
Db 340 VIDLSKEGGETYLGDERGECKNGIQAKTSLSEKSTINVSKEKGGRAIWDGIALI 399
QY 486 --NLSTOKI-----FLNSGDENFENVELLSKEQ-----NINPLTLKESQSHL 530
Db 400 DGNINAGSGDIAKYGFEVETSGHDLPFKDNAIVDAKEMLDQDNVSIHAETAGRS----- 455
QY 531 PDGNLSHFEGYGDWTFESWKSDEGSHLIAMTPKNVYPHPEROSTLVANTLM-----N 584
Db 456 ---NTSEDEDTG-----SGNSAS---TPKR-----NKEKTLNTLESILKKG 494
QY 585 TYSDMAA-----VOSMINTAMGAYLF--GTWGSAYV---SNLTFYADS-----SK 626
Db 495 TFEVITANQRIYVNSSIN-LSNGSLTLMSEGRSGGVEIINDITDGTGANLTIYSG 553
QY 627 PIDNMHRSGLYLFGLSTHSLDHSFCLAAGOLL-----GKSS 664
Db 554 WDVHKNISLGAQGNINITAKODIAFEKGSNOVITIGOGTITSGNOKGFFRNVSJLNGTGS 613
QY 665 DSFIETSTETSYIATVQAOIATSL--MKISAQACYNESIHLELTKYRSFSKGFSGSWH- 720
Db 614 GLOFTTKRNTKRYATNKKEGTINISGVNISMVLPKNEGYD-KFKGRY-----WNL 665
QY 721 --SVAVSGEVCASIPVNSGCLFSSFSIYSKLOGFSGTODGFEBSGCEIRFSASSTR 777
Db 666 TSLNVESEGEF--NLITDSRGSAGLITLOPYNLNGISFNKD---TTFVVERNARVNF- 718
QY 778 NISLPIGI 785
Db 719 DIKAPIGI 726

RESULT 12
US-08-469-880-2
Sequence 2, Application US/08469880
Patent No. 5876733
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995

```

Db 76 VVH---GTATMOVDGNKTTIRNSVD---AIIWKKOPNDQNEWVOLFQEN---NNSAVF 125
QY 180 AQTFSLSRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFTGNSATNGAICICISDL 239
Db 126 NRVTS-----NOLSKGLDLSNGQVFL-----INP-----NGITIGDAII 162
QY 240 NTSEKGSLSIACNOETLPASN---SAKEKGAIYAKH--MWLRNGPVYFSTINSAKIGG 293
Db 163 NTNGFTASTLDISNENIKARNFTFEOTKDALAEIVNHGLITVGKGSVNLIGKVKYKNG 222
QY 294 AIAIOSGGSLSIAGSGSVLPFNNSORTSDQGLVRNAIYL-----EKDAI-LSSLEARNG 347
Db 223 VLSV-NGGSISLLAG-----OKITISDIINPTIYYSIAAPENEAVALGDIPIAKGG 271
QY 348 DILFDPPIVQESSKESPLSSLSQASVTSPTATASPLVYIOTSANRVSFSSERLSEEEK 407
Db 272 NT-----NVRRAATIRNGKLSADSV 291
QY 408 TPDNLTSQLQPIELKSGRLVLKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISANOQAAGKGLMITGDKVTLKTGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTIHAP-----485
Db 340 VIDLSGKEGGETYLGDEREGCKNGIOLAKKTSLEKGSTINVGKKEGKRAIYMGDIALI 399
QY 486 --NLSTOKI-----FLNSGDENFEYENVLLSKEQ-----NNIPLTLSEKQSHLH 530
Db 400 DGNINAGSGDLAKTGTFVETSGHDLFIKDNALYDAKEWLDLDPNVSINAEIATGRS-----455
QY 531 PGDNLSSHFGYGDWTFWSKMDSEGHSLANMTPKNVVPHPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTLNTLESILKKG 494
QY 585 TYSDMA-----VOSMINTIAHGAVL--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TPNVNTIARORIYVNSSTIN-LSNGSLTLWSEGSNGGVEIINDITGDDRGANLTIYSGG 553
QY 627 PIDMHRSLGYLFGISTHSLDHSFLAAGQL-----GKSS 664
Db 554 WVDVHNINISLGAGNINITAKODIAFEKGSNOYITGOGTITSGNOKGFRRNNVSLGTS 613
QY 665 DSFTISTETTYATVQAOIATSL--MKISAQACYNESIHLELKYRSEKSGEFGSMH- 720
Db 614 GLOFTTKRNTKAYAITMKFEGLINISGVNISMVLPKKNESGYD-KFKGRTY-----WNL 665
QY 721 ---SVAVSEGCASPIYVSGSLFSSSFTSKLQGFSGTQDGFESSSGEIRSFASSPR 777
Db 666 TSLNVSESEF--NLFTIDSGSDSAGTLTPYVNLNGISFNKD---TFVVERNARVNF- 718
QY 778 NISLPIGI 785
Db 719 DIKAPIGI 726

```

```

RESULT 14
US-08-617-697-2
Sequence 2, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

```

```

Query Match      3.0%  Score 134.5, DB 2, Length 1536:
Best Local Similarity 19.4%: Pred. No. 0.0058:
Matches 153: Conservative 123: Mismatches 253: Indels 259: Gaps 38:

```

```

QY 120 IIHQKGOI-SLRNNGSMSCRNHAEBSGGAISADAFSLQHNLTFTFEENSSKNGCAIQ 179
Db 76 VVH---GTATMOVDGNKTTIRNSVD---AIIWKKOPNDQNEWVOLFQEN---NNSAVF 125
QY 180 AQTFSLSRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFTGNSATNGAICICISDL 239
Db 126 NRVTS-----NOLSKGLDLSNGQVFL-----INP-----NGITIGDAII 162
QY 240 NTSEKGSLSIACNOETLPASN---SAKEKGAIYAKH--MWLRNGPVYFSTINSAKIGG 293
Db 163 NTNGFTASTLDISNENIKARNFTFEOTKDALAEIVNHGLITVGKGSVNLIGKVKYKNG 222
QY 294 AIAIOSGGSLSIAGSGSVLPFNNSORTSDQGLVRNAIYL-----EKDAI-LSSLEARNG 347
Db 223 VLSV-NGGSISLLAG-----OKITISDIINPTIYYSIAAPENEAVALGDIPIAKGG 271
QY 348 DILFDPPIVQESSKESPLSSLSQASVTSPTATASPLVYIOTSANRVSFSSERLSEEEK 407
Db 272 NT-----NVRRAATIRNGKLSADSV 291
QY 408 TPDNLTSQLQPIELKSGRLVLKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISANOQAAGKGLMITGDKVTLKTGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTIHAP-----485
Db 340 VIDLSGKEGGETYLGDEREGCKNGIOLAKKTSLEKGSTINVGKKEGKRAIYMGDIALI 399
QY 486 --NLSTOKI-----FLNSGDENFEYENVLLSKEQ-----NNIPLTLSEKQSHLH 530
Db 400 DGNINAGSGDLAKTGTFVETSGHDLFIKDNALYDAKEWLDLDPNVSINAEIATGRS-----455
QY 531 PGDNLSSHFGYGDWTFWSKMDSEGHSLANMTPKNVVPHPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTLTLNTLESILKKG 494
QY 585 TYSDMA-----VOSMINTIAHGAVL--GTWGSAY--SNLFYAHDS-----SGK 626

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Gitschler, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 8: FROM 1 TO 2319
US-08-212-133A-8

Query Match 2.9%; Score 132.5; DB 1; Length 2319;
Best Local Similarity 20.2%; Pred. No. 0.02;
Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FEDCT-----MESLPAL-----CANASODPLVIGNSYCWVSKLHTDPKEALF 92
DB 775 FDDSTIPKNDMEIEPOFEIEIAEMLKQVSYSVSDMLMLGQSHR-TPHGLFLSDGGEALY 833
QY 93 K-----EKGDLSIONFRLFTDCSSKSSPSIHKNGQSLRNNGSMSC 139
DB 834 EAIHDDSPNAIDSNQSPKVTQLR-----PESHSEKIVFTPOPLGRNKSLSL--- 883
QY 140 RNHAESSGAISDAFSLQHNITFT-----AFEE-----NSKNGGAI 178
DB 884 ETTIEKMKKLGQVSSLSPSNLTITLSDNLKATEKTDSSGCFPMVHSSSKLSLTJAR 943
QY 179 QAOFTLSRNVSPISFARNADLNGAICCSNLICSGNVNPLFFTGNSATNGAICISD 238
DB 944 GKRAYSLVGSHPVPLNSSEUSD-----SNLDS-----TLWYSGESLPRDMLSTEND 991
QY 239 LNTSEK--GSLSLACNOETLLFASNSAKKEGCAIY---AKHMLRYNCPVSFINNSAKIGG 293
DB 992 RLREKRFGIALTLKDNFTLFKDNVSLMKTKYNNHSTTEKLTHTESPTIENSTDLQD 1051
QY 294 AI-----AIOSSGSLSLAGEGSLVFNNSQRTSDQGLVNAIYLEDALSL-SLEARNG 347
DB 1052 AILKVNSEID--EYVALIHGDTLCKNST-----YLRNLNHLNRTSTKKN 1095
QY 348 DILF---FPIVQESSKESPLSSLOASVTSPTPATASPLVIQTSANRSVIFSSERLSE 404

DB 1096 DIFHRKDEDPIDQDEENTIMPSKMLFLS-----ESSMNEKKTNGNSL-----NSE 1142
QY 405 EEXTDNLTLSQLOP1ELKSGRLVLDRAVLISAP---SLSDPOLALLMEAGTSIKTSDD 461
DB 1143 QHSPROLYLYL-----FKRYVKNOSFLEKKNVYVEODGFTKNGLKDMAPPHNS 1194
QY 462 LKLTATSLPLHS---LDTEKSV--TIHAPNLSTQKIFLSN---SGDENFENVELLSKE 512
DB 1195 IFLTTIS-NVHENGKRNDEKNIOEIEKALIEKAVYLPQVHEATGSKNPLKDLILGTR 1253
QY 513 QN-----NIPLLTLSKEQSHLPLPDGNLSSHFQYGDWTFKSKDSDECHSLIA 560
DB 1254 QNLSLEYHVHVLQNTTSINNSTNTVOIHM-----EHFFKRR-----KDKETNSEGLV 1301
QY 561 NMTPKNYVHPEROSTLVANTIMNTYSDMQAVOSMINTIARHGAYLFGTWSAVSNLPTA 620
DB 1302 NKTRE-----MVKN-----YPSQKNITTORSKRALQOFRISTOW---LKTINCS 1342
QY 621 HDSGKPIDNMHRSGLYFGISTHSLDHSFLAAGOLLGKSSDSFISTETTSYIATV 680
DB 1343 TOCIIKQID--HSEKMKR--ITKSSLSDSVYIKSTTQ--NSSDSHYKTSAPFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHLEKTR 707
DB 1397 RSPFQNKFSHYOA---SSYIYDFKTK 1419

RESULT 17
US-08-474-503-6
Sequence 6, Application US/08474503
Patent No. 5744446
GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,503
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMU106CIP(3)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6500
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.

QY 140 RNHAEGSGAISAQAFSDHNYLFT-----AFEE-----NSSKNGCAI 178
DB 884 ETTIEVKKKKLGLVSSLSPLNMTTILSDNLKATFEKTDSSGFPDMVHSSSKLSTTAF 943
QY 179 QAQTFSLRNVPISFARNRADLNGAICCSNLICSGVNNPLFTGNSATNGAICICSD 238
DB 944 GKAAISLVGSINPLNASENSD-----SNLDS-----TLMYQESLPRONILSIEND 991
QY 239 LNTSEK--GSLSLACNOETLFPASNAKEKGAIY--AKHMYLRNGVPVFINNSAKIGG 293
DB 992 RLREKRFHGIALTKDNTLFDKNVSLMKTNTKYNHSTTNEKLHSTESPSTIENSTDLDD 1051
QY 294 AT-----AIOGSGSLILAGEGSLVFONNSQRTSDOGLVRNAILYEKDAIIS-SLEARNG 347
DB 1052 AILKYNSEIQ--EYVALHHDGTLGKNST-----YLRNHLNRTTSTKTK 1095
QY 348 DILF---FDPPIVOESSKESPLPSLSLOASVTSPTPATASPLVLOTANSRVSISSERLSE 404
DB 1096 DIFHKKDDPIPODEENTIMPSKMLFLS-----ESSNMFKKTNGNNSL-----NSE 1142
QY 405 EKPDPNLTSQLQPIELKSGRLVLDRAVLAP---SLSDPQALLINEAGTSLKTSDD 461
DB 1143 QEHSPKQOLVLYM-----FKKYVKNOSFLSEKNKVTYBODGFTKNIGLKDMAFPHNMS 1194
QY 462 LKATLSTPLHS---LDTEKSV--TIHAPNLSIOKIFLSN---SGDENFVENNELLSKE 512
DB 1195 IFLTLLS-NVHENGHNQKNOEIEKEKALIEEKVYVLPDVHEATGSKFLKDILILGR 1253
QY 513 QN-----NIPLLTSKEOSHLPLPDGNLSSHFGYGDVTFWSKDSDEGSLIA 560
DB 1254 QNISLXYHVAVPLQVITSIINNSTNVQIHM-----EHFKNR-----KDKETNSEGLV 1301
QY 561 NMTPKNVYHPERKOSTLVANTIMNTYSDMAVQSMINTIAGGAILFCWGSAVSNLFA 620
DB 1302 NKTR-----WVKN-----YPSQKNITTOQRKRALQOFRUSTOW--LKTINCS 1342
QY 621 HDSGKPLDNMHRSLGVLFGISTHSLDHSFLCGLAGOLGKSSDSFISTETTSIATV 680
DB 1343 TOCIIKQID--HSEKMKR--ITKSSLSDSVYIKSTOT--NSSDSHYKTAFAFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHELTK 707
DB 1397 RSPFQKRFHVQA---SSYIYDFKTK 1419

RESULT 19
US-09-037-601-6
Sequence 6, Application US/09037601
Patent No. 6180371
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/037, 601
FILING DATE: 26-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Lachich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA
Patent No. 6180371
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
US-09-037-601-6

Query Match 2.9% Score 132.5 DB 4: Length 2319;
Best Local Similarity 20.2% Pred. No. 0.02; Indels 187; Gaps 35;
Matches 151; Conservative 109; Mismatches 300;
QY 46 FECDT-----MELFPAL-----CAHASODPLVYLGNSYCWFSKLTITDREALE 92
DB 775 FKDSTIPKNDMEKIEQFEIEAEMLKVQSVSDMLMLGQSHP-TPHGLFLSDQGEALY 833
QY 93 K-----EKGLDLSIONFRELSTDCSSKESPSIITHQNGQLSRNNGSMFC 139
DB 834 EAIHDHSPNAIDSMCEGPKVTQLR-----PESHSEKIVFTPOPGIQLRSKSL-- 883
QY 140 RNHAEGSGAISAQAFSDHNYLFT-----AFEE-----NSSKNGCAI 178
DB 884 ETTIEVKKKKLGLVSSLSPLNMTTILSDNLKATFEKTDSSGFPDMVHSSSKLSTTAF 943
QY 179 QAQTFSLRNVPISFARNRADLNGAICCSNLICSGVNNPLFTGNSATNGAICICSD 238
DB 944 GKAAISLVGSINPLNASENSD-----SNLDS-----TLMYQESLPRONILSIEND 991
QY 239 LNTSEK--GSLSLACNOETLFPASNAKEKGAIY--AKHMYLRNGVPVFINNSAKIGG 293
DB 992 RLREKRFHGIALTKDNTLFDKNVSLMKTNTKYNHSTTNEKLHSTESPSTIENSTDLDD 1051
QY 294 AT-----AIOGSGSLILAGEGSLVFONNSQRTSDOGLVRNAILYEKDAIIS-SLEARNG 347
DB 1052 AILKYNSEIQ--EYVALHHDGTLGKNST-----YLRNHLNRTTSTKTK 1095
QY 348 DILF---FDPPIVOESSKESPLPSLSLOASVTSPTPATASPLVLOTANSRVSISSERLSE 404
DB 1096 DIFHKKDDPIPODEENTIMPSKMLFLS-----ESSNMFKKTNGNNSL-----NSE 1142
QY 405 EKPDPNLTSQLQPIELKSGRLVLDRAVLAP---SLSDPQALLINEAGTSLKTSDD 461
DB 1143 QEHSPKQOLVLYM-----FKKYVKNOSFLSEKNKVTYBODGFTKNIGLKDMAFPHNMS 1194

QY 551 DSDEGSLIANTPKNYVHPEROSTLVANTLMNTYSDMVAOSMINTIAHGAYLFTCM 610
DB 665 --EEG-----KN-----NGNLANTFKGSKSEONR--LTTGTNLNGDL 698
QY 611 GSAVNLFEVAHDSCKPIDNMHRSGLYFGISTSLDHSCLAGOLLGSSDSFTIS 670
DB 699 KVEKTELL--SGRPT--HARDIA--GISSTKKDOH--FANNVVEDDOMINN 746
QY 671 TETSYIATVOAL--ATSLMKISAQ--ACYNESIHE-----LTKYRSF----- 711
DB 747 FKATINNTNATLYSGRVANITSNITASDNKAVHIGKAGDTVCVRSDYGYVCTTD 806
QY 712 --SKGPGSMHVAASGEVCAISPIVNSGGLFSSFSIJSKLOGSCODDGEES-- 765
DB 807 KLSDALNSFNATVNSGVNLS--GNANFVLKANLFGTISGTGNSOVLRTENSHML 862
QY 766 -GEIRSFASFRNISLPIG--ITFEKKSQKTRY-----YFELGAYIIDL 808
DB 863 TGD-----SNVNOJLQKCHILHNAQNDANKVTYNTLTVSLSGNSGF--YLTDL 912
QY 809 KRDVESGPVLLKANV--SWDAPMANLDSRAVMERLTNORALHRLQTLNLVSCVLRGQSHS 867
DB 913 S-NKOGDKVYVTKSATGNETLOVADKTGPTKNEILLFDASHATRNMLNVSIV--G 965
QY 868 YSLDLG 873
DB 966 NTVDLG 971

RESULT 24
US-09-356-952-4
Sequence 4, Application US/09356952
Patent No. 6117663
GENERAL INFORMATION:
APPLICANT: Borjack-Sjodin, Ann
APPLICANT: Margarit, S. M.
APPLICANT: Bor-Sogi, Dafna
APPLICANT: Cole, Philip
APPLICANT: Kuriyan, John
TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
FILE REFERENCE: 600-1-228N
CURRENT APPLICATION NUMBER: US/09/356,952
CURRENT FILING DATE: 1999-07-19
EARLIER APPLICATION NUMBER: 60/093,631
EARLIER FILING DATE: 1998-07-21
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1589
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-356-952-4

Query Match 2.8%; Score 126; DB 4; Length 1589;
Best Local Similarity 18.1%; Pred. No. 0.04;
Matches 162; Conservative 141; Mismatches 322; Indels 272; Gaps 37;

QY 112 SSKESSPSIIHONGLOSLRNNGSMFCRNHAGSGCAISADA--FSLQHYVLTARENS 170
DB 143 SSXSSSRSSLN-----SLGNSAYLVHPKPNPSSRRGSSSTLSASLSVAH-----AETSS 191
QY 171 SKNGGAIOAOTFSLSRNVPISFARNRADLNGAICCSNLSGNNVPLFTGNSATNG 230
DB 192 GHANTVSMNNSPFSAPNDASHIT--POSSNFNMSALISQDMTKSAD-----GSSEMT 242
QY 231 GAICCSIDLTSKSGSLSLACNOETLFASNSAKEKGALYAKHMV-----LRYNCP 281
DB 243 NAI-----NNNNE-----TNLOT-----SGEKAGPPLVAEETIKILPLEIEMIING- 284
QY 282 VSTINNSAKIGALAIQSGSLSLAGEGSVLVRQNNNSQRTSDQIVNAYLYLKD----- 336

DB 285 --IRSNIASTWSPILIT-----KTSQKLY--YYNKDLIDYC 318
QY 337 ---AIISSLEARNGOILFFDP-----IVQSSSKESPSPSLQASVTSPPPATAS 383
DB 319 SELPLISNIMESDIOCESEPFPPNDHLVNTYTRDLRKNANIEDSSSTRKSESEONNS 378
QY 384 PLVIOTSANRVSIFSERISEEKTDPNLTISOLOO--PELKSGLVYLKRAVLASAPIS 441
DB 379 SLMEKQSKETEDGNNSINDDNNNKNEFNAGBSLNS-----LSAPDLT 428
QY 442 ODPQALLIMEACTSLKTS-----DLKATLSPLHSLDTEKSVYIHAENLSIOKIF 493
DB 429 QNIGSRVAVPSRSSLSLAKSDIFYHVSROIKLMT--ELQDLVYVYTKAH-----KMF 478
QY 494 LNSGDENYENVELLSKE-----QNNIPLITLSKE-----Q 525
DB 479 LKEN--RLNFTKRYFDLSDSIVFTOLGCRMLQHEIRAKSCSEIKKIFKGLISSLSRISIN 537
QY 526 SHLHLPDGNLSSHFGQDWFPMKQSD-----EG-----HSLIA----- 560
DB 538 SHLY-----FDSAFIRKKMDTMNDKDNQENNCSTBEDCKIEYDSVHDLVSVPLSGK 592
QY 561 -----NMT-----PKNYVHPEROSTLVANTLMNTYS-- 587
DB 593 RNVSTSTDTLTPMRSSFVNEENDENFVJGPRRSVSVYTPRTSIGNSTLEDFSPSN 652
QY 588 ----DMQAVOSMINTIANGATLFGTWSAVSNLYAHDSGKPIDNMHRSGLYF--G 641
DB 653 KNFKSAKSIYEMVDV-----EFSKFLRHVOLLVFLVQSVFSDNTLPQLLPREFKGS 705
QY 642 ISTHSLDHSFCLANQOLKSSDSFTSTENT-----SYIATVOALATSLMKISAQ 695
DB 706 FSGGSWTNPSTFTIDEFGNATKNKAVTSNEVYASSKSSISLIPPKADAIASAGYS 765
QY 696 CYNESIHLELTK-----YRSFSEKGFSGHSAVAGEVCASIPVNSGGLF--SS 744
DB 766 ANSETNSOJDLAKASSAASSVFTFPRPNHRTFSARAKSKKKKPLVLDLNTMKKS 825
QY 745 FSTFSKLOGFCGTODGFEESGGEIRFSASFRNISLPIGITFEKKSQKTRYVFLGAY 804
DB 826 SOIFEKLNMTG-----EHLKIISKPRIRINLEIN--STYEQINQ-----NVL 868
QY 805 IODIAKRDVESGPVLLKANVSNAPMANLDSRAVMERLTNORALHRLQTLNLVSCVL 861
DB 869 LLEILENLDLSTFINKLNIKTIPSTILDLSEEF-----VHAMSSVSVL 914

RESULT 25
US-08-687-956A-23
Sequence 23, Application US/08687956A
Patent No. 5861157
GENERAL INFORMATION:
APPLICANT: BURRILE, JAMES P
APPLICANT: MATTHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
FLOOR
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996

XX 26-APR-1999 (first entry)
 XX Chlamydia pneumoniae surface exposed protein Omp10.
 DE
 XX
 XX Omp10; outer membrane protein 10; surface exposed protein;
 KW antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
 OS Chlamydia pneumoniae.
 XX
 XX W09858953-A2.
 PN
 PD 30-DEC-1998.
 XX
 XX 19-JUN-1998; 98WO-DK00266.
 PF
 PR 23-JUN-1997; 97DK-0000744.
 XX
 XX (BIRK/) BIRKELUND S.
 PA (CHRI/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 DR WPI: 1999-105610/09.
 DR N-PSDB: X06822.
 XX
 XX Species-specific test for identifying mammals infected with
 PT Chlamydia pneumoniae - comprises detecting antibodies specific for
 PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
 PT these proteins
 XX
 XX Claim 7: Page 60-62; 115pp: English.
 XX
 CC This polypeptide comprises the novel 98.4 kDa surface exposed
 CC protein Omp10 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06822) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W8816-27), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used
 CC in the diagnosis of C. pneumoniae infection in mammals. The
 CC nucleic acids and proteins can also be used in the immunization of
 CC mammals, the nucleic acids being particularly useful as DNA
 CC vaccines for effecting in vivo expression of antigens. The
 CC vaccines may also prevent atherosclerosis and bronchial asthma,
 CC which are possibly associated with C. pneumoniae.

Query Match 20.7%; Score 940.5; DB 20; Length 928;
 Best Local Similarity 31.5%; Pred. No. 2,8e-70;
 Matches 287; Conservative 136; Mismatches 338; Indels 149; Gaps 32;
 OY 85 TDPKALFEKGDLSIONFR-LSTFTDCSSKESVPSIIHOKNG-OLSLRN----- 132
 DB 52 tnaadgtltnlqdvstlnagsptalcscfketgnlsgfyqfllqndaganctft 111
 OY 133 -----NGSMVF-----CRNHAEGSGAISADAFSLQHNLYLTAFEESSKNGGAI 178
 DB 112 ntaankllstfgsfyislqtnatcgalkscgacsiqny-sclyfgnlsndngal 170
 OY 179 QAOTFSLRNVPISFARNRADLNGAICCSNLICSGN-VNPLFTGNSATN----- 229
 DB 171 qgsstslslnnp-lfaknkakqkqalyslqgllnnlnlaasentlaanggalyle 229
 OY 230 -----GGAICCTSDLNTSEKGSLSLACNOETLFASSAKK 265

DB 230 assfssnkaistfinsvatsatgalyccs--tsapkvltltdngelnfintait 287
 OY 266 GGAIVAKHMYLVNPGVPSINNS-----AKIGAIATIGSGSLSLAGSVLPONNS-- 318
 DB 288 ggalyldnlnvissgpcllfkmsaldlaapiggaladagslslsalygdlfngntv 347
 OY 319 -ORTSDOGLVRNAYL-EKDAIILSLIARNCD-LEFDPVGESSKESPLDSSLQASVT 375
 DB 348 kgasssqtlrnslnlgnlnakivqlraeqnltlyydpdltlnht-----aals 396
 OY 376 SPTPATASPLVIQTSANRVSIFSSERLSEDEKT-PRNLTSQLOQPELKSGLRVLKDRV 434
 DB 397 dalnlngpdlagpnaygqvtlsvgeklscseaaadnlkltlqgpltlagqslskgyt 456
 OY 435 LSNPSLSODPQALLMEAGSLKTSDDLKATLSPILHSLDPEKSTVIAHPNLSIQIFL 494
 DB 457 lvaakstsqspyscllndagcltletadqcltnnlvlnvdsllketkcltkatqas-qtlvtl 515
 OY 495 SNS-----GDENFEYENVVLISKDQNN---IPLTLSEKQ-SHLPLPDGNLS----- 537
 DB 516 sgslslvdpsgnvyedv-----swmpqvfcsltlcaddpanlhld--laadpleknpl 568
 OY 538 HPGYGDWTFEWMKDSDEGSLIA--NWTPKNVVHPEROSTLVANTLNTYSMDQVQSA 595
 DB 569 hmgvgqmwalswgdeatkakaatlwtcltgnpnpertgltvaantlwsfvdvrsiqpl 628
 OY 596 INTIANGAVLFGTWSAVNLFYANDSSGKPRIDNMHHRSLGYLFGISTHSDHSCFLA 655
 DB 629 valkvrsqelrtglwceglisnft-hkdscklnkgrfnhsagyyvgatcltlaadnltaa 686
 OY 656 AGOLLKSSDPSFPTSTFETTYIATVQAO-LAT-----SIAMK-----ISAQACYN 698
 DB 687 fccglfgkdrhflnkraasyaaslhqlatlspsllylpgseseqpvltlaqsl 746
 OY 699 ESIHELTKYRSFSEKGEFGMSHVAVSGEVCASTPIYS-NGSGIFSSFSIFSLQGSST 757
 DB 747 ysktmtklytqapk-gesswyndgcslajasslphta]sheglfhayfflikweasyih 805
 OY 758 ODGEF-SSGEIRFSSASFRNLSLPIGTPEKKSOKTRTYRYVLAGAYIODLKRDESSE 816
 DB 806 qdsfkerntllvrsidsqdlinvsvpigliiterfstrerasyaevatlvyadvyrknpdct 865
 OY 817 VVLIKNVSDMPANLDSRA-----YMFRLTNORALHRLQTLNLVNSCVLRGQSHSY 868
 DB 866 talllnlswtlctlnlsrgeqigragllyafs-----pnlvtenlsmeltgsrsry 918
 OY 869 SLDLGTTYRF 878
 DB 919 nadlgykqf 928
 RESULT 6
 Y35054
 ID Y35054 standard: Protein; 927 AA.
 XX
 AC Y35054:
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed polypeptide.
 XX
 KW respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 OS Chlamydia pneumoniae.
 XX
 PN W09927105-A2.
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-1B01890.
 XX

Query Match	20.3%	Score 919.5;	DB 20;	Length 930;
Best Local Similarity	31.0%	Pred. No. 1.7e-68;		
Matches 260;	Conservative 150;	Mismatches 343;	Indels 87;	Gaps 26;

QY	98	LSIQRNRLFSF-----TDCSSKESSPSIIHOKNGOLSLRNNSSPFCK---NHAEGSGGA	149
Db	119	lftfgnslsilaapglctvasgkts-----sagalnltctgttllfsgnvsneannmga	174
QY	150	ISADAVSLQHNLYFTAFEESSSKGNGAI-QAOTFSLRNWSPISFARNADLNGAICC	208
Db	175	lctkltslsgntslstlftssnakkllgalaissaasltgqglvfmnkgctggga1--	232
QY	209	SNLICGNV---NPLFTFGNSATN-----GGALICSLDLNTSEKGLSLUACQETLFAFMS	261
Db	233	-gfeasssltqnssllfsgntldaagkxgalya---ektgetrplltlsgnksltfne	288
QY	262	AKERGCATYKKNHWLRYNGVFSFIN-----SAKIGALIAIGSGSLIAGEGVLVFN	316
Db	289	svtggagichgjdlaeaqgltsmrcnqtaagkxgalatadsqslsaangdltlg	348
QY	317	NS-QRTSDQGLVKNATYLEKDALLSSLLEARNG-DLFLPPIQESSMSNPSPSLQSV	374
Db	349	nltltsapetrlnatlylgsakltlrlraagqsllyfyoblsnctgadvl-----	400
QY	375	TSPPAPASPLVIOTSNRNVIFSSERLS-EEKKRPDNLITSOLOOPTILKSGRLAKRA	433
Db	401	tlmpdpsnsp1-----dysgvlvtsgetkslaadeakaadntsltkpqlaagslalxgnv	456
QY	434	VLSAPSLSDOPALLIMEAGTSLKTSDD-LKLATLSPLHSIDTEKSVYIHP-----N	486
Db	457	eldvngftqtegsellmpgkpkadteaalsklxlvddalsalegnksvsietaeanlct	516
QY	487	LSIQKIFLNSGDNFENYENNELSKQNNIPLTLTLESKQSHLTPDGNLS-----HF	539
Db	517	ltsplvfdqssg--nfyeshlindaftqplrvlvtataaasady1--dalltspqvcrephy	573
QY	540	GYGDMWFMSKMDSECHSLIANMTPKRVNPHPEROSTLVANLTMTNYSMDQAVQSNINI	599
Db	574	ygqglweatwadtclaksgltwltclynprpetraaavpds1wasfdlrltqglmtsq	633
QY	600	AHGCATVLEGTGSAVSNLFYAHDSGKPRIDMNHNRSLGYLFGISTHSLDHSFCUACOL	659
Db	634	ansltyqgtfvaasgtanflh-kdksygn-qafnhtksygyivggsaadfsenitsvafcl	691
QY	660	LCKSSDSFTTSTETSYATYVQAQ-----LATSLMK-----ISAOACTNES	700
Db	692	fgkdddlflventshnylasy1lqhraf1g9lpmrpfsgsltdmkdip1lmaeq1sytyc	751
QY	701	HEHLTKTRSPKBEFGGSMHVAVSGEVCASIP1-VNSNGCLTFSSPFIKSLQFGSGTOD	759
Db	752	kndmdrtfisy-pegagswtmnsgal1elgslaly1lkeapdfggyrpfllkfgavysrq	810
QY	760	GFEESSGEIRSPASSPFINSLPIGITEFKKSOKTRTYVYFLGAY1ODLKRDESGPVYL	819
Db	811	nftesgaearatddgd1vncslpvgltlrekliseedekmfelslayldvrykmptrstsl	870
QY	820	LKNAVSWDAPMANLDSRAYMFLTNQRLH-RLQTLNVSCLVSLGSGHSYSLDGTYRF	878
Db	871	mvsqaswtlcklnlarqatlaasagshltlspvelsgaayaeltsgashlynvdcglrlysf	930
RESULT	8		
ID	Y90240	standard; Protein: 930 AA.	
XX	XX	Y90240:	
XX	XX	29-AUG-2000 (first entry)	
XX	XX	Chlamydia antigen CPN100639.	
KW		Chlamydia antigen; diagnosis: infection; community acquired pneumonia	

[illegible]

PR 20-AUG-1998: 98US-0097195.
 PR 20-AUG-1998: 98US-0097196.
 PR 20-AUG-1998: 98US-0097197.
 PR 27-AUG-1998: 98US-0097191.
 PR 17-AUG-1999: 99US-0376770.
 XX (CONN-) CONNAUGHT LAB LTD.
 PA Murdin AD, Oomen RP.
 PI WPI: 2000-224703/19.
 DR N-PSDB: 261509.
 XX Novel antigens and corresponding DNA molecules that can be used to
 PT prevent, treat and diagnose disease caused by Chlamydia infection in
 PT mammals, especially humans.
 XX
 PS Claim 19; Fig 15-E: 201pp; English.
 CC Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides
 CC are present in the bacterial membrane structure, in the external
 CC vicinity of the membrane structure, in the inclusion membrane
 CC structure, in the external vicinity of the inclusion membrane
 CC and in the cytoplasm of the infected cell. The polypeptides may be
 CC used to prevent, treat and detect the presence of Chlamydia infection
 CC and/or the presence of Chlamydia in a sample. The polypeptides may
 CC also be used to induce an immune response in a mammal. The vaccine
 CC vector comprising the poly nucleotides is used to induce an immune
 CC response in a mammal. Antibodies directed against the polypeptides
 CC may also be used therapeutically to treat and/or prevent a Chlamydia
 CC infection.
 CC
 SQ Sequence 918 AA:
 Query Match 19.1%; Score 865; DB 21; Length 918;
 Best Local Similarity 31.2%; Pred. No. 6.6e-64;
 Matches 253; Conservative 138; Mismatches 326; Indels 94; Gaps 29;
 QY 120 IHHKNGQLSLRNNGNSFCRNHAEHGGGGAISADAFSLQHNHYLTFAFENSSKNGCAIQ 179
 DB 150 lvfesignldnnesse-----nggainktklstltsgrtfaflgnsssgg99a1y 201
 QY 180 AQRTS-USRNPSPISFARNRADLNGAICCC-SNLCGNNVPLPPTG-NSATNGCAICCI 236
 DB 202 asgsvisehngllstfgnsatstsgaaisaegnlvisnngn-iffdgckatnggaic- 259
 QY 237 SDLMTSEKSLSLACNOETLFASSNAKEGAIYAKHNVIRY-NGPVSFINNSA---KI 291
 DB 260 nkagandpplilistgnesihflmtagnsga1yckk1vlsgrgvalifsmnkaeatpk 319
 QY 292 GGAIAIGSGSLISLAGEGSLVLPQNNNSORT--SDOGLVRNAIYLEKDALISLEARNND- 348
 DB 320 ggalaildsgeisladsadignilfegntstgspasvtrnaidlsnakflinlratrpk 379
 QY 349 ILFPPIYOESSKESPLPSLOASVTSPPTPATAPLVLOTGANS-----VIF 397
 DB 380 vlfypl-----lssgatklslnxadagsgnllygy1vf 414
 QY 398 SEERISEEE-KTPDNLTLOOPIELKSGRLVLDRAVLASPSLQDPQALLIMEAGTSL 456
 DB 415 sgekliseelkpkpnhkfstfgave1aaga1vldgvtvvanlticvegskvmdgltf 474
 QY 457 KTSSD-LKLATLSTPLHSD-TEKSY---THAPNLSIOKIFLSNGDENFENVELLSK 511
 DB 475 eesaeagvltng1a1nds1dgnkalikaaskaakdvalsgimlvdaagnyehnm--s 532
 QY 512 EGNMNPILTLTSMK--QSHMLPDPG---NLSHFYGGDMTFEWSKWDDEGSLTA--NMP 564
 DB 533 gqvfp1l1elsagqmtldtp1l1ctlnhygqgnm1vvdadataktacltwck 592
 QY 565 KNYVPEROSTLVANTLNTYSDMOAVOSMINTIAHGAYLFGTWSAVSNLFYAHDS 624

DB 593 tgyknpnpergpllvpsn1vgstfvdrslqslmdrstssjssstnlwsg1ad-f1hbdqk 651
 QY 625 GRPIDNMHRRSGYLFEGTSTHSLDHSFCLAGCOLLKGSSDSFISTETTYIATYQAO- 683
 DB 652 gmq-rsyhnsagyal99gfflaseenlfnafctqldh1vaknhbhvayagasyrh 710
 QY 684 --LATSIMKI-----SAQACVNESIHETKTKYRSFKEGFGSMHSAVANGEVIC 729
 DB 711 lgeskl1ak1lsgnsd1p1vfnar1ayghldnm1tky1gypvk-gawgnadfg1ecg 769
 QY 730 ASIP1VSNQ-SGLFSSFSIFSKLQGFSGTQDGEESGGEIRSFASFRN1SLP1G1TFE 788
 DB 770 galpvasgr1swdch1p1nlem1yahbndf1kengtegr1fsged1fn1avp1g1kte 829
 QY 789 KKSQKTRTYIFELGAY1ODLKRDVEGSPVVLKNVSMDBAN1DSRAVMEFLTNQRL 848
 DB 830 kfcdks-tyd1st-ayvpd1rindpgcttl1mvsqds1wstc1g1sarqal1vra1ghnal 887
 QY 849 -HRLQTL1ANVSCVLRGQSHS1SLD1GTYRF 878
 DB 888 asnfefvsg1evelfrgss1ryald1g9rf1gf 918
 RESULT 13
 ID W88421 standard: Protein: 928 AA.
 W88421:
 AC W88421:
 XX 26-APR-1999 (first entry)
 DT
 XX Chlamydia pneumoniae surface exposed protein Omp8.
 DE
 XX Omp8; outer membrane protein 8; surface exposed protein; antigen;
 KM infection; diagnosis; vaccine; atherosclerosis; asthna.
 KW
 XX Chlamydia pneumoniae.
 OS
 XX W09858953-A2.
 PN
 PD 30-DEC-1998.
 XX
 PF 19-JUN-1998: 98WO-DK00266.
 XX
 PR 23-JUN-1997: 97DK-0000744.
 XX
 PA (BIRK/) BIRKELUND S.
 PA (CHR1/) CHRISTIANSEN G.
 XX
 PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
 PI Mygind P;
 XX
 DR WPI: 1999-105610/09.
 DB N-PSDB: X06820.
 XX
 XX Claim 7; Page 53-55; 115pp; English.
 CC This polypeptide comprises the novel 90.0 kDa surface exposed
 CC protein Omp8 of the human respiratory pathogen Chlamydia
 CC pneumoniae. Its amino acid sequence was deduced from DNA (see
 CC X06820) isolated from a C. pneumoniae expression library. The
 CC invention provides 12 novel surface exposed proteins, Omp4-Omp15
 CC (see W88417-28), and nucleic acid sequences encoding them (see
 CC X06816-27). A new species specific test is claimed that is used
 CC to identify mammals (including humans) infected with Chlamydia
 CC pneumoniae. The test comprises detecting antibodies specific for
 CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
 CC membrane proteins, especially by PCR. The proteins are also used

```

OY 128 LSLRNNGSMFCRNHAEGSGAIS-ADAFSLOHNLFTAFEESSKNGCAIO---AQTF 183
DB 175 llyqiknsaifqntskkxggaistgyltlemjltikfnekvatggaldgaastf 234
OY 184 SLRNVSPISFARNRAD---LNGCAICCS-NLIGSGNVNPLEFTGNSATNGAICISDL 239
DB 235 tanhe---lifsqntksngaangaincsgdltfndtslllqenstmgdgaic----- 286
OY 240 NTSEKGSLSLACNOETLFASNAKEKGAIYAAHM-VLRINGEVSFINN---SARIGCA 294
DB 287 ---stglstisgdsainvigtsgkxggaistaalikllyggagallsnvthaltprl9ga 343
OY 295 IAOOGSGSLILAGESVLPONNSQRTSDGLV--RNAIYLEKDAIISLEARNCD-ILF 351
DB 344 lftngqsgslftcggqdlvteggvtttapatktrnvlnheslaktglaaaggaalyf 403
OY 352 FDPVIOESSKESPLPSSLOAVTSPPATASPLVIQTSANR---SVIFSSERLSEEEK 407
DB 404 ydpi-----tndtgaadnlfrinevsanqklsysivfsgerlstaee 445
OY 408 TPNLITSLQOPIELKSGRLVLDRAVLASPSLQDPAQLLIMAGTSKTS--DLKAT 466
DB 446 laenltarlnqpvllvsgslvlgvlllqglfsgqestllldlglstlkastledvln 505
OY 467 LSLPLHSLDTEKSVTIAH---PMLSTIOKILPLNSGDENFENELUSKEONNIPLAT-- 520
DB 506 lsnadliygnpnlvsaankliltglalvnaadgaelyenhltdsgdsvtkxlpq 565
OY 521 ----LSKEOSHLHLPDGNLSHFYGODMTFSMKDSEGHSLIAN--WPKNVPIPER 573
DB 566 aggtlftqdaaqkplvavapsrphlygqghmvqviptgtpsqanleavrtgylpmpet 625
OY 574 QSLTAVNTMTYSDMOAVQSMINTIAHGAVLCTGWSANVNLFYAIDSSGKPRDMNH 633
DB 626 qgslypnalwgsfvdqraigeimwnsqllcqezyvwagqian-llhcdkline--hgyrth 682
OY 634 RSLCYLPFISHSDDHFCFLAAGLAKSSDSPTSTETTSYIATV----- 680
DB 683 sgvgylvgvthafsdalinalfcqlfscrkdyvsknhglsygsvfltdletfrpgq 742
OY 681 ----QAOLATSLMKISAOCYNESIHETKTRYSFSEKGEFSMHSVAVGEVCASIDP 734
DB 743 fytssaseacnqvvtlimqslsyrnmndmcklctly-pegagswandvfgjefatly 801
OY 735 VSNNGSLFSSIFSKLQCFSTOGFEFSSEIFSSFSASFRNLSLPRGIFFEKKSQKT 794
DB 802 ypnstflldyyspflrlqctyahgedfketgevrhftsgdldfnlavpivgktertsdck 861
OY 795 RITYYFFLAYIOD-LKRDVEGPPVLLKNAVSDAPMANLDSRAVFRLTNORALH-RLQ 852
DB 862 rgsyeltlayvdpvltkdpks--tatlagaatwstghnlsrtgqlrltgnhcllnpgle 919
OY 853 TLLNVCVLRGQSHSYSLDLCTYRF 878
DB 920 vfhsgaieirgsrnylnlggkyrf 945

```

```

RESULT 15
ID Y90237 standard; Protein: 928 AA.
XX Y90237;
XX 29-AUG-2000 (first entry)
XX Chlamydia antigen CPN100635.
XX Chlamydia antigen CPN100635.
XX Chlamydia antigen: diagnosis: infection: community acquired pneumonia:
XX therapy: upper respiratory tract disease: bronchitis: sinusitis:
XX asthmatic bronchitis: adult-onset asthma: acute exacerbations of asthma.
XX Chlamydia pneumoniae.
XX

```

```

FH Key Location/Qualifiers
FT Peptide 1..43
FT Protein /note="signal peptide"
FT Protein 44..928
FT Protein /note="mature CPN100635"
MO200032794-A2.
PD 08-JUN-2000.
PE 01-DEC-1999: 99MO-CA01147.
XX
XX 01-DEC-1998: 98US-0110339.
XX 01-DEC-1998: 98US-0110340.
XX 01-DEC-1998: 98US-0110427.
XX 01-DEC-1998: 98US-0110428.
XX 01-DEC-1998: 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murlin AD, Oomen RP, Wang J:
XX WPI: 2000-412339/35.
XX DR N-PSDB: A30849; A30850.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma.
XX
XX Claim 16: Fig 3: 174pp: English.
XX
XX This sequence is a Chlamydia antigen of the invention, designated
XX CPN100635. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to treat infections. The
XX expression of the proteins and may be used to produce the protein antigens they
XX nucleic acids may also be used to produce the protein antigens they
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX (for example) community acquired pneumonia, upper respiratory tract
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 928 AA:
XX
XX Query Match 18.7%; Score 847; DB 21; Length 928;
XX Best Local Similarity 30.1%; Pred No. 2.3e-62;
XX Matches 284; Conservative 141; Mismatches 409; Indels 108; Gaps 31;

```

```

OY 12 CAIISSTAVLFGODPL--GETALLTKPNHNVV----TFEEDCTNESL--FPALCAHAS 63
DB 20 cscvfaataenipsdsfsgstntgtytknttgdtyltlqngysaalekyctf 79
OY 64 QD--DPIYVLGNSYCMFVSKLHITDPEKALFKKGDLSIONFRELSTDOSSKKE--SSPS 119
DB 80 sdtesistagkxys--lsflnkssaeq-----aalsvtdknlsitlsgfslflfaas 132
OY 120 I-----HOKNGOISLRNNGSMFCRNHAEGSGAISADAFSLOHNLFTAFEEENS 170
DB 133 svltlpsgkxavxcgalditdmgtlftkqyceenggaistknlsknstgisstfgnk 192
OY 171 SKG---NGCAIOAO-TFSLRNVSPISFARNRADLNGAI-CCSNLIGSGNVNPLEFTGN 225
DB 193 ssatgkkggaicactgtvdcntnaptlfnmiaeaggaistncttgnfs-lyfsen 251

```

OY 505 NVELLSKEQN-----NIPLLTJSKEQSHLHPDGNLSHFYGGDWTFSWKD-----SDEGH 556
DB 531 nhd-lygktdqdfsvqjsaigtatcttdvpavrpvaerphlygygvtgmctwvddtaetpkx 589
OY 557 SLIAWMTKPNVYHPHROSTLAVNTIAMNTYSDMAOVASQINTIAHGAYLFTQWGSANVN 616
DB 590 fatlwtntgylpmpgrgqplvpnsalwgsfsgldqglqiersalltlcsdrgfwaagvan 649
OY 617 LFYADSSCKPIDNMHNSLGLFGLSTHSLDHSCLAGOLCKSSDSFTSTSTTSY 676
DB 650 -fldkdkkgek-rykyhksgyaiggaqtcsehlisafcgfsgdkxfivaknhtdy 707
OY 677 IATVQAQLATSL-----MKISAQACYNESIHBLKTKYRSFSEKGG 717
DB 708 agafyqhhtecsgfigclldklpgawshkplviegqlayshvsnldkkytaty-pevxyg 766
OY 718 SMHSVAVSGEVCASIPIVNSGSLFESSFIFSKLQGFSGTQDGFERSGGEIRFSASFR 777
DB 767 swgnafmmiigasshsypelyhcfityapyiklnltlytqdsfsekgtgrsfddsnlf 826
OY 778 NISLPIGITEFEKKSQKTRTYVFLGAYIDLKRDVESGPVYLKNAVSDAPMANLDSRA 837
DB 827 nlslpgrvtefctsdndfsgldtlysvpllrndpcttalvisgswetyannlartga 886
OY 838 YMER 841
DB 887 lgyr 890

RESULT 17
Y90238
Y90238 standard. Protein: 885 AA.
AC Y90238;
XX
DT 29-AUG-2000 (first entry)
XX
DE Mature Chlamydia antigen CPN100635.
XX
KM Chlamydia antigen; diagnosis: infection; community acquired pneumonia;
KW Chlamydia: upper respiratory tract disease; bronchitis; sinusitis;
KM asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.
XX
OS Chlamydia pneumoniae.
PN
XX W0200032794-A2.
XX
PD 08-JUN-2000.
XX
PE 01-DEC-1999; 99WO-CA01147.
XX
PR 01-DEC-1998; 98US-0110339.
PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Mordin AD, Oomen RP, Wang J;
XX
DR WPI: 2000-412339/35.
DR N-PSDB: A30849, A30850.
XX
PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma.
PS Claim 16; Fig 3; 174pp; English.
CC This sequence is a Chlamydia antigen of the invention, designated
CC CPN100635. The nucleic acids (and their complementary sequences) may be

CC used as diagnostic agents for detecting the presence of nucleic acids
CC encoding Chlamydia antigens in samples according to standard methods,
CC and therefore, for diagnosing Chlamydia infections. For example, they may
CC be used as primers and probes for diagnostic polymerase chain reaction
CC (PCR) assays. Antisense sequences may be used to down regulate
CC expression of the proteins and may be used to treat infections. The
CC nucleic acids may also be used to produce the protein antigens they
CC encode according to standard recombinant DNA methodologies. The
CC proteins may then be used as antigens for the production of antibodies
CC (i.e. as vaccines) for preventing infection by Chlamydia. The
CC antibodies may also be used as diagnostic reagents for detecting
CC infections. Chlamydia is a pathogen implicated in the development of
CC (for example) community acquired pneumonia, upper respiratory tract
CC disease (especially bronchitis and sinusitis, asthmatic bronchitis,
CC adult-onset asthma and acute exacerbations of asthma in adults.
XX

Sequence 885 AA:

Query Match 18.7%; Score 846; DB 21; Length 885;
Best Local Similarity 30.5%; Pred. No. 2,5e-62;
Matches 279; Conservative 135; Mismatches 382; Indels 120; Gaps 30;

OY 28 LGETALLTKPNINIVCTFFEDCTMESLFPALCAHASDDPLVLYGNSYCFWVKLHTDP 87
DB 25 lgsdaalkg-----c--fsdt-----eslsfagkyrs--isflnkss 60
OY 88 KEALFKKKGDSIONFFRLFTDCSKE--SSPSII-----HOKNQLSLRNNGSM 136
DB 61 aeg-----aalsvtcdknslstgtslftlaapsvltlpsgkavvcgdlftdngtl 115
OY 137 SFCRNHAEGSGAISADAFSLQHNLYLTAPEENSSK-----NGAIOAO-TPSLRNVSPI 192
DB 116 lfxqdyceenggaistknlslkntsgisfegnkssatgkkgalcatyrdtlnnapt 175
OY 193 SFARNRADLNGAL-CCSNLCISGNVNPFFTGNSAT-----NGAICISDNTSEKSL 247
DB 176 lfsmlaeaaggaalnstgcttgnts-lvisensvatalaagngal-----sgdadv 226
OY 248 SLACNOETLFAFSNSAKEKGATVAKHMYLRVNGPVS--FINN-----SAKIGATAIQSG 300
DB 227 llsqngsvtfisngqavangaalylakklilaagggggnpfsnmlyqgtlaagngasllaa 286
OY 301 GSLSLAGEGYLFONNS--QRTSDGLVIRNAIYLEKDAIILSSLEARNG-DILFDPPIVOE 358
DB 287 gecsifseagchlyngnaltatpqtikrnsldtsgtkheltlrajshtfitypitan 346
OY 359 SSSKESPLPSSLOASVTSPTPATASPLVIQTSANRSVIFSSPRLSEEE-KTPDNLTSOLO 417
DB 347 laadstdtlnlnkadagn-----stlsgslyfsgklsedeakvadnltstlk 395
OY 418 QPTELKSGRLVKKRAVYAPSLSDOPALILMEAGTSILKTS-DUKLATISIPLSHDT 476
DB 396 gpvtlaagnlvlykrgvltldtkgftqtaagssvlimdgltkateevltlgslpdslyge 455
OY 477 EKSVTIHAP-----NLSTOKIFLSNSGDENFENVVELLSKEQN-----NIPLLTJSKEOS 526
DB 456 gkkvylaasaaknvalsgp11ldng--nayeend-lygktdqdfsvqjsaigtatcttd 512
OY 527 HHLHPDGNLSHFYGGDWTFSWKD-----SDEGHSLIAWMTKPNVYHPHROSTLAVNTL 582
DB 513 vpavrpvaerphlygygvtgmctwvddtaetpkktaltawltgylpmpgrgqplvpnsl 572
OY 583 WNTYSDMAOVASQINTIAHGAYLFTQWGSANVNLFYADSSCKPIDNMHNSLGLFGL 642
DB 573 wgsfsgldqglqiersalltlcsdrgfwaagvan-fldkdkkgek-rykyhksgyaig 630
OY 643 STRHSDHSECLAGOLCKSSDSFTSTSTTSYIATVQAQLATSL----- 688
DB 631 aagtcsehlisafcgfsgdkxfivaknhtdyagafighhtecsgfigclldklpgs 690
OY 689 -----MKISAQACYNESIHBLKTKYRSFSEKGGSMHSVAVSGEVCASIPIVNSGSLFS 743

XX	30-DEC-1998.
XX	PD
XX	PF
XX	19-JUN-1998; 98WO-DK00266.
XX	PR
XX	23-JUN-1997; 97DK-0000744.
XX	(BIRK/) BIRKELUND S.
PA	(CHRI/) CHRISTIANSEN G.
XX	Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI	Mygind P;
XX	WPI: 1999-105610/09.
DR	N-PSDB: X06816.
XX	Claim 7; Page 40-42; 115pp; English.
XX	This polypeptide comprises the novel 98.9 kDa surface exposed
CC	protein Omp4 of the human respiratory pathogen Chlamydia
CC	pneumoniae. Its amino acid sequence was deduced from DNA (see
CC	X06816) isolated from a C. pneumoniae expression library. The
CC	invention provides 12 novel surface exposed proteins, Omp4-Omp15
CC	(see W88A17-28), and nucleic acid sequences encoding them (see
CC	X06816-27). A new species specific test is claimed that is used
CC	to identify mammals (including humans) infected with Chlamydia
CC	pneumoniae. The test comprises detecting antibodies specific for
CC	Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC	membrane proteins, especially by PCR. The proteins are also used
CC	in the diagnosis of C. pneumoniae infection in mammals. The
CC	nucleic acids and proteins can also be used in the immunization of
CC	mammals, the nucleic acids being particularly useful as DNA
CC	vaccines for effecting in vivo expression of antigens. The
CC	vaccines may also prevent atherosclerosis and bronchial asthma,
CC	which are possibly associated with C. pneumoniae.
XX	Sequence 928 AA:
SQ	
	Query Match 18.6%; Score 845; DB 20; Length 928;
	Rest Local Similarity 28.8%; Pred. No. 3.3e-62;
	Matches 247; Conservative 148; Mismatches 345; Indels 118; Gaps 26;
QY	97 DLISIONRFLSTDCSSKESSPS-IIHOKNGOLS-----LRNCGSMFCRNHAEGSGGA 149
Db	: : : : : : : : :
	113 ntfgsfllstf-----dssptstvtgtgtllsaggyvlnenikrlvvagnfstadgga 166
QY	150 ISADAFSLQHNLFTAFENSSKGNGGATQAOTFS--LSRNVSPISFARNRADLNGGAIACC 208
Db	: : : :
	167 ikgasfiltqsgdalifsnssstkkgaiattagarianntgyvfslniastsgaidd 226
QY	209 SNLICSGVNPJFFTGNSA-TNGGATCCISDLNTSEKGS--LSLACNOPTLPFASNAKEK 265
Db	: : : :
	227 egtslisnnklyfegnaaktggalc-----ntkasgspeliisnknktlifasnvaets 281
QY	266 GGAIYAKHMVLRYNGVPVSFINNSAKI----GGAIAIOSGGSLILAGESVLFQNNSQRT 321
Db	: : :
	282 ggaiahakkalalsgggftefirnnvssatpkgsaidasgeilsaetgnitfvrrtltt 341
QY	322 --SDOGLVRNAILYLEKDALLSILE-ARNGDILFFEDPVOESSSKESPUPSSLQASVTSP 378
Db	: : : :
	342 tgstdtprknaignsgkfteleaknhnhtiffdydpitsegtssdvklknngsagalnp 401
QY	379 PATASPPLVIQTSAANRSVFISSERLSEEE-KTPPNLTSQLQQPTIELKSGRIVLKRAVL 437
Db	: : : :
	402 qgt-----ilfsgetldeikvadnklksftqpvsisggkilkqkvtyles 448
QY	438 PSLSQDPQALLINEAGTSLTKTSS-DLKLATFLSIPLHSLDTEKSVTHAPNWSITQKIFLS 496

CC	which are possibly associated with C. pneumoniae.									
XX	Sequence	918 AA;								
SQ	Query Match	18.6%; Score 843;	DB 20;	Length 918;						
	Best Local Similarity	31.2%; Pred. NO. 4.8e-62;								
	Matches	249;	Conservative 137;	Mismatches 325;	Indels 86;	Gaps 28;				
QY	134	GSMFCRHAEGSGAISADAFSLQHNVLFTAFENSSKNGGAIQAOTFS-LSRNVSP	192							
DB	156	gnldqnenassengaintklsitgrfvaflgnssqggggaigasvdsisenagil	215							
QY	193	SFARNRDLNGGAIACC-SNLICSGNVNPLFTFG-NSATNGAICISDLNTEKGSLSLA	250							
DB	216	sfgnssattsagaisaegnvlisnnqn-iffdgckattnggaide-nkaganpdpillts	273							
QY	251	CNQTFLFASNAKEKGGAIAKHMVLRV-NGPVSPFINNSA-----KICGAIAIQSGSLSI	305							
DB	274	gneslhfnntagnsggaitytkkvlssgrgvgvifsnkkaanatkpggaialidsgeisi	333							
QY	306	LAGEGVLFQNNRSORT--SDQGLVNAIYLEKDAILSSLEARNGD-ILFEDPIVOESSK	362							
DB	334	sadlnilfegnttstspasvtrnaidlasnkflnratrgnkvlfydpl-----	386							
QY	363	ESPLPSSLQASVTSPATASPLVIQTSANRS-----VIFSSPRISEEP-KTPD	410							
DB	387	-----tssgatklsinkadagsgntyegyivfsggeklseee.lkxpd	428							
QY	411	NITSLOQPILKGRVLKDRAVLSAPLSODPQALLIMEAGTSLKTSDD-LKLATLSI	469							
DB	429	nikstftgavalaagalkvdkgtvvtantitqevskvymdggttfeasaegvtnglai	488							
QY	470	PLHSLD-TEKSV--TIHAPNLISIOKIFUSGDEFYENVLLSKSQNNIPLLTSLKS-	524							
DB	489	nldslgtdnkalkikaasakvaigspimlvdaqngnyehnl--sqgvfplielasag	546							
QY	525	-OSHLHLPDQ---NLSSHFYOGQDWTFSWKDSDGHSLIA--NWTPKNYVPHPEROSTLV	578							
DB	547	tmttdipdptlntttnhygggtgllvvvdataaktknatllwtktg/kpnergpglv	606							
QY	579	ANTLWNTYSDMQAVQMINTIAHGGAYLFTGWSAVSNLFYAHDSCKPIDNMHHRSIGY	638							
DB	607	pnsalwgsfvdvrsiqslmrdstsslsstnllwsgiad-flhedqkgnq-rsyrrhsag	664							
QY	639	LFGISHTSLDDHSCFAAGOLCKSSDSFTTSTTTSYIATVQAO---LATSLMKI---	691							
DB	665	alggrfteseninfafcdqfgydkdhvaknhtviegamsyrhlgesktlakilsgn	724							
QY	692	-----SAQACYNESIHELKTYRFSKRGFCSMHSAVSGEVCASIPVNSG-SGLF	742							
DB	725	sdslpfvfnarfayghtdnmttkygsypk-gswgndafglecggaipvvasgrsw	783							
QY	743	SPSIFSKLQSGTQDCPDESSCEIRSPSASSFRNLSDIGITFEKKSOKRTYYFLG	802							
DB	784	dthtppfnlemyahqndfkengtegrfsqsdlnlavpvgikfekfsdks-tydlisi-	841							
QY	803	AYTODLARDVESGPVVLKKNVSDNAPMANLDSRAYMFLTNQRAL-HRLQTLNNSCVL	861							
DB	842	ayvpdvirndpgctttlmvsgdsvstcgtslsrqallivregnhhafasnfefvsqfevel	901							
QY	862	RGOSHYSLDLGTYYRF	878							
DB	902	rgsrsyaidlggrgfg	918							
RESULT	22									
W88428										
ID	W88428	standard; Protein; 945 AA.								
AC	W88428:									
XX										
DT	26-APR-1999	(first entry)								

XX	Chlamydia pneumoniae surface exposed protein Omp15.									
DE	Omp15; outer membrane protein 15; surface exposed protein; antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.									
XX	Chlamydia pneumoniae.									
OS	Chlamydia pneumoniae.									
PN	WO9858953-A2.									
XX	30-DEC-1998.									
PD	19-JUN-1998; 98WO-DK00266.									
XX	23-JUN-1997; 97DK-0000744.									
PR	(BIRK/) BIRKELUND S.									
XX	(CHRI/) CHRISTIANSEN G.									
PA	Birkelund S, Christiansen G, Knudsen K, Madsen A;									
PI	Mygind P;									
XX	WPI; 1999-105610/09.									
DR	N-PSDB; X06827.									
XX	Species-specific test for identifying mammals infected with Chlamydia pneumoniae - comprises detecting antibodies specific for outer membrane proteins of C. pneumoniae or nucleic acids encoding these proteins									
PT	Claim 7; Page 71-73; 115pp; English.									
XX	This polypeptide comprises the novel surface exposed protein Omp15 of the human respiratory pathogen Chlamydia pneumoniae. Its amino acid sequence was deduced from DNA (see X06827) isolated from a C. pneumoniae expression library. The invention provides 12 novel surface exposed proteins, Omp4-Omp15 (see W88417-28), and nucleic acid sequences encoding them (see X06816-27). A new species specific test is claimed that is used to identify mammals (including humans) infected with Chlamydia pneumoniae. The test comprises detecting antibodies specific for Omp4-Omp15 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of C. pneumoniae infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with C. pneumoniae.									
XX	Sequence	945 AA;								
SQ	Query Match	18.2%; Score 825;	DB 20;	Length 945;						
	Best Local Similarity	28.2%; Pred. NO. 1.7e-60;								
	Matches	244;	Conservative 153;	Mismatches 371;	Indels 98;	Gaps 25;				
QY	74	SYCMFVSKLHITDPKEALFKKGDGDL-----SIOFRLFSTDCSSKESPSIIHOKNGO	127							
DB	117	sftdfslvitespkpsavstgkgsilvssgavqlqdlntlvitsasvedgvi--kgnsc	174							
QY	128	LSLRNCSMSFCRHAEGSGAIS-ADAFSLQHNVLFTAFENSSKNGGAIQ---AOTF	183							
DB	175	liqgiknsaifgqntskkkggaistctgtltnlgtlkfneknkavissggaiddgaastf	234							
QY	184	SLSRNVSPISFARNRAD---LNGGAICCS-NLICSGNVNPLFTGNSATNGAICCSIDL	239							
DB	235	tanhe---lifsqnktsgnaangaincsgdlftdntslilqenstmqdgalc-----	286							
QY	240	NTSEKGSLSLACNQTFLFASNAKEKGGAIAKHM-VLRYNGPVSPFINN-----SAKIGA	294							
DB	287	---stgttisgdsdinvgitsgdkggaisaaskilggggalifsnvnnvthatplgga	343							

OY 655 AGQLLGKSSDFTSTSTTSYATVOALATSLMKIS-----AAQYNESIHELTK 707
DB 742 gfgqltkskdylvghghsnvfatvysnithkslfgsffggtservtysrnekvks 801
OY 708 YRSFSGEGFSGSHVAVSGEVCASIPVNGSGL-FSSFSIFSKLQGFSGTODGFEESSG 766
DB 802 ytklpk-grcswnncwlgelnlpitlssrllnkqilpfvkaevayathgqgentp 860
OY 767 EIRSFASFRNLSIPGITFEKKSQKTRTYFFLGYIQDLKRDVSGPVLLKNVSW 826
DB 861 egrifghhllnvavpvgvrfgknsnrpdytliivayapdvyrhnpdcdtllpingatw 920
OY 827 DAPMANLDSRAYMRLTNORALHR-LQTLNVSCVLRGSHSYSLDLGTTTYRF 878
DB 921 tsignlstrllvqasshslsvndvleifghcgcdlrrtsarkytlldigsklrf 973

RESULT 24
Y37231
ID Y37231 standard; Protein; 182 AA.
AC Y37231;
XX
XX
DT 07-OCT-1999 (first entry)
XX
XX Chlamydia trachomatis cellular envelope protein.
DE
XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
XX Chlamydia trachomatis.
OS
XX
XX W09928475-A2.
XX
XX 10-JUN-1999.
PD
XX
XX 27-NOV-1998; 98WO-IB01939.
XX
XX 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
XX (GEST) GENSET.
PA
XX
XX Griffais R;
XX
XX WPI; 1999-371125/31.
DR
XX
XX Genome sequence of Chlamydia trachomatis
PT
XX
XX Disclosure: Page 992; 1755pp; English.

XX Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
CC Chlamydia trachomatis (see 201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perinephritis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
XX
XX Sequence 182 AA;

Query Match 17.5%; Score 794; DB 20; Length 182;
Best Local Similarity 97.5%; Pred. No. 4.le-59;
Matches 157; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 705 KTKYRSFSGEGFSGSHVAVSGEVCASIPVNGSGLFSSFSIFSKLQGFSGTODGFEESS 764
DB 1 knidgsfsgksgwhsvavsgvecasipvngsglffsfkslqgfsqtdgfees 60
OY 765 SGEIRSFASFRNLSIPGITFEKKSQKTRTYFFLGYIQDLKRDVSGPVLLKNV 824
DB 61 sgeirsfasfrnlsipgltfeksksqtrtyyfigayiqdkrdvsgpvllknv 120
OY 825 SWDAPMANLDSRAYMRLTNORALHRLQTLNVSCVLRGOS 865
DB 121 swdapmanldsrmyrltnqralhrlqtltnvscvlrgos 161

RESULT 25
Y99843
ID Y99843 standard; Protein; 925 AA.
XX
XX Y99843;
XX
XX 15-SEP-2000 (first entry)
XX
XX Chlamydia pneumoniae processed 98 kDa outer membrane protein CPN100640.
DE
XX Chlamydia; 98 kDa outer membrane protein; antigen; immunogen; infection;
KW vaccine; antibacterial; community acquired pneumonia; bronchitis;
KW sinusitis; acute respiratory disease; upper respiratory tract disease;
KW asthma; atherosclerosis.
XX
XX Chlamydia pneumoniae.
OS
XX
XX W0200032784-A1.
DN
XX
XX 08-JUN-2000.
PD
XX
XX 01-DEC-1999; 99WO-CA01148.
XX
XX 01-DEC-1998; 98US-0110439.
PR 03-MAY-1999; 99US-0132272.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
PI
XX
XX WPI: 2000-412330/35.
DR N-PSDB: A48838, A48839.
XX
XX New polynucleotide encoding the Chlamydia 98 kilodalton outer membrane
PT protein, useful for preventing or treating Chlamydia infection -
XX
XX Claim 16; Fig 1; 98pp; English.

XX The present sequence is the mature, processed form of CPN100640, the
CC 98 kDa outer membrane protein from Chlamydia pneumoniae. Chlamydia
CC pneumoniae is a common cause of community acquired pneumonia and upper
CC respiratory tract symptoms and diseases, including bronchitis and
CC sinusitis. It also has an association with atherosclerosis and asthma.
CC The 98 kDa outer membrane protein is a C. pneumoniae-specific antigen
CC which can confer immune protection against chlamydial infection. The
CC nucleotide sequence encoding the protein or the protein itself may be
CC administered as a vaccine to prevent or treat infection and they may
CC also be used to diagnose infection. The gene encoding CPN100640 was
CC amplified from Chlamydia pneumoniae genomic DNA by PCR.
XX
XX
XX Sequence 925 AA;

Query Match 17.1%; Score 773; DB 21; Length 925;
Best Local Similarity 28.0%; Pred. No. 4.le-56;
Matches 236; Conservative 143; Mismatches 378; Indels 86; Gaps 22;

OY 97 DLSTQNERFLSFDTDCSSKSSPS--IIHQKNGQLSLRNGSMFCRNHAECSGGAISADA 154
DB 108 nllfnfdrslscpslllspgtgcalksvgnlslgnsqilftqnfssdnggvntkn 167

OM of: US-09-677-752-4 to: N_Geneseq_0401.* out_format : pfs
Date: May 6, 2001 8:37 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+P2n.model -DEV=xlp
-O=/cgn2_1/USP70_spool/US09677752/runat_04052001_132443_4961/app_query.fasta_1.948
-DB=N_Geneseq_0401 -QFMT=fastcap -SUFFIX=mg -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -START=1 -MATRIX=blosum62
-DELOP=6.000 -DELEXT=7.000 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09677752 -CGN1_1_255 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-677-752-4
Query length: 878
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 148.120000

score_list:

Sequence	Strd Orig	zscore	Escore	Len	! Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:A64751		4513.00	6882.52	0.0	2643
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:201425		4395.00	6632.08	0.0	1038602
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64757		4373.00	6668.22	0.0	2601
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:CB1914		947.50	1365.81	3.9e-68	273254
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30852		944.50	1414.69	7.3e-71	2784
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30851		944.50	1414.02	8.0e-71	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		940.50	1408.55	1.6e-70	2787
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X91990		930.00	1321.46	1.1e-65	1230025
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30853		912.00	1364.79	4.4e-68	3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30823		906.50	1356.43	1.3e-67	2793
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06817		883.00	1320.34	1.3e-65	2815
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		870.00	1300.53	1.7e-64	2790
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		866.00	1293.55	4.1e-64	3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:CB1914		866.00	1240.94	3.5e-61	273254
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30852		855.00	1292.19	4.9e-64	2957
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06828		857.00	1280.62	2.2e-63	2787
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30854		853.50	1274.40	4.8e-63	3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		851.00	1270.38	8.0e-63	3050
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		849.50	1268.47	1.0e-62	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		849.50	1267.52	1.2e-62	3200
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		847.00	1264.64	1.7e-62	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30847		844.50	1261.60	2.5e-62	2757
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06827		825.00	1231.38	1.2e-60	2838
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A27342		811.00	1208.72	2.2e-59	3150
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30848		810.00	1208.63	2.2e-59	2784
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30848		786.00	1171.86	2.5e-57	2784
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48838		773.00	1150.88	3.6e-56	3050
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A47458		763.50	1136.43	2.3e-55	3021
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06539		759.50	1125.83	9.0e-55	4435
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64752		757.50	1127.52	7.1e-54	2949
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06556		753.00	1119.13	2.1e-54	3354
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06557		749.00	1113.10	4.6e-54	3324
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A50035		735.00	1093.05	6.1e-53	2950
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06818		735.00	1092.65	6.4e-53	3052
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48839		717.00	1066.04	1.9e-51	2808
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28691		708.50	1054.26	8.8e-51	2526
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28690		708.50	1053.26	9.9e-51	2750
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:X06819		705.50	1049.66	1.6e-50	2526
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28709		574.50	847.49	2.9e-39	2865
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28708		574.50	846.38	3.3e-39	3150

/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28715 + 558.50 822.69 6.9e-38 2937
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28714 + 558.50 821.87 7.7e-38 3150
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28713 + 536.50 789.55 4.9e-36 2796
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A28712 + 536.50 788.73 5.4e-36 3000
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:261586 + 532.00 781.65 1.3e-35 3050

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A64751

seq_documentation_block:

ID_ A64751 standard; DNA; 2643 bp.
XX AC A64751;
XX DT 02-FEB-2001 (first entry)
XX C. trachomatis pmpI gene coding sequence.
XX Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial; ss.
XX Chlamydia trachomatis.
XX WO200034483-A2.
XX PD 15-JUN-2000.
XX PF 08-DEC-1999; 99WO-US29012.
XX PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX (CORI-) CORIXA CORP.
XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
WPI: 2000-431303/37.
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection
XX comprises immunogenic portion of Chlamydia antigen, which comprises
XX amino acid sequence encoded by polynucleotide sequence
XX Claim 1: Pages 171-172; 256pp; English.
XX The present invention relates to new nucleic acid sequences and the
XX proteins encoded by the nucleic acid sequences. The encoded proteins
XX comprise an immunogenic portion of a Chlamydia antigen. The encoded
XX proteins are useful for the serodiagnosis and treatment of Chlamydia
XX infection. Chlamydiae are intracellular bacterial pathogens that are
XX responsible for a wide variety of human infections. C. trachomatis
XX infection is one of the most common sexually transmitted diseases and can
XX lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
XX and infertility. Trachoma due to ocular infection with C. trachomatis is
XX the leading cause of preventable blindness worldwide. C. pneumoniae is a
XX major cause of acute respiratory tract infections in humans and is also
XX thought to play a role in the pathogenesis of atherosclerosis and
XX coronary heart disease. The present sequence is a nucleic acid sequence
XX isolated in the present invention.

Sequence 2643 BP; 702 A; 651 C; 503 G; 786 T; 1 other;

alignment_scores:

Quality: 4513.00 Length: 878
Ratio: 5.158 Gaps: 0
Percent Similarity: 99.658 Percent Identity: 99.544

alignment_block:

US-09-677-752-4 x A64751

601 HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLe 617
 1807 CACGGAGGCGCTATCTATTGGAACTGGGATCTCTGCTTTCTAATTT 1856
 617 uPheTyrAlaHisAspSerGlyLysProIleAspAsnTrpHisHis 634
 1857 ATTCATGTTCCACGACGCTCTGGAAACCTATCGATAATTGGCATCAT 1906
 634 rGserLeuGlyTyrLeuPheGlyLysSerThrHisSerLeuAspHis 650
 1907 GAAGCCTTGGCTACCTATTGGTATCATCTACTACAGCTTTAGATGACCAT 1956
 651 SerPheCysLeuAlaGlyGlnLeuGlyLysSerSerAspSerPh 667
 1957 TCITTCCTGCTTGGCTGAGGACAAATTAATCGGAAATCGTCGGATTCCTT 2006
 667 eileThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnL 684
 2007 TATTACGCTACACAAACGACCTCTATATAGCTACTGTACAAACGCAAC 2056
 684 euAlaThrSerLeuMetLysLysSerAlaGlnAlaCysTyrAsnGluSer 700
 2057 TCGTACCTCTCTAATGAAATCTCTGCACAGGCATGCTACAATGAAAT 2106
 701 IleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheG 717
 2107 ATCCATGAGCTAAACAAATATTCGCTCTCTCTTAAGAAGGATTCCGG 2156
 717 ySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProI 734
 2157 ATCTGGCATAGCTGTCAGTATCCGAGAGAGTGTGCGCATCTGCTTCT 2206
 734 leValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerLys 750
 2207 TTGTATCAATGGTTCGGAGCTGTTCAGCTCTCTCTCTATTTCTCTAAA 2256
 751 LeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyG 767
 2257 CTGAAGGATTTTCAGGAACACAGAGCGGTTTCAGGAGAGTTCGGGAGA 2306
 767 uIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleG 784
 2307 CATTCGGTCTCTTTCTGCCACTCTTTCCAGAAATATTTCACTTCTCTAT 2356
 784 lYileThrPheGluLysLysSerGlnLysThrArgThrTyrTyrPhe 800
 2357 GAATAACATTTGAAAAAATCCCAAAAAACACGAACCTACTATTACTTT 2406
 801 LeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyProVa 817
 2407 CTAGAGCCTACATCCAGACCTTGAACGCTGATGGAATCGGACCTGT 2456
 817 lValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuA 834
 2457 AGTGTACTCAAAATAGCCGCTCTCTGGATGCTCCTATGGCGAATCTGG 2506
 834 spSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArg 850
 2507 ATTCAGAGCCTACATGTTCCGGCTTACGAATCAAGAGCTCTACACAGA 2556
 851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSe 867
 2557 CTTACAGAGCCTGTTAAATGTCTGTGTGCTGCTGGGCGAAAGCCATAG 2606
 867 rTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
 2607 TTACTCCCTGGATCTGGGAGCACCATTACAGGTTT 2640

seq_name: /SIDS2/gcdata/geneseq/geneseq/NA1999.DAT:201425
 seq_documentation_block:
 ID 201425 standard; DNA; 1038602 BP.
 XX

Z01425;
 07-OCT-1999 (first entry)
 Complete genome sequence of Chlamydia trachomatis.
 Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
 nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 Chlamydia trachomatis.
 W09928475-A2.
 10-JUN-1999.
 27-NOV-1998; 98WO-1801939.
 04-NOV-1998; 98US-0107077.
 28-NOV-1997; 97FR-0015041.
 17-DEC-1997; 97FR-0016034.
 (GEST) GENSET.
 Griffais R;
 WPI; 1999-371125/31.
 Genome sequence of Chlamydia trachomatis
 Claim 1; Page 373-656; 1755pp; English.
 The present sequence represents the complete genome of Chlamydia trachomatis. Open reading frames (ORFs) of the genome encode polypeptides X36754-X37949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis, pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.
 Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment_scores:
 Quality: 4395.00 Length: 876
 Ratio: 5.052 Gaps: 2
 Percent Similarity: 99.315 Percent Identity: 99.087
 alignment_block:
 US-09-677-752-4 x Z01425
 Align seg 1/1 to: 201425 from: 1 to: 1038602
 1 MetArgProAspHisMetAsnPheCysCysLeuCysAlaAlaIleLeuSe 17
 471863 ATGGACCTGATCATATGAACCTTCGTGCTATGCTGCTGCTATTTGTC 471912
 17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGluThrAlaLeuL 34
 471913 ATCCACACGGTCTCTTTGGCCAGGATCCCTTAGTGAACCGCCCTCC 471962
 34 euThrLysAsnProAsnHisValValCysThrPhePheGluAspCysThr 50
 471963 TCACATAAAATCCTAATCATGCTGCTGTACATTTTTCAGGAGCTGACC 472012
 51 MetGluSerLeuPheProAlaLeuCysAlaHisAlaSerGlnAspAsp.P 67
 XXX

```

667 heileThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGln 683
|||||
473862 TTATTAGCTACAGAAACGACCTCTTATATAGTACTGTACAGCGCAA 473911
|||||
684 LeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSe 700
|||||
473912 CTCGCTACCTCTCTAATGAAATCTCGACAGCGCATGCTACAATGAA 473961
|||||
700 rIleHisGluLeuLysThrLysTyrArg_SerPheSerLysGluGlyPhe 716
|||||
473962 TATCCATGAGCTAAACAAATATCGCTCTCTCTAAAGAGGATTC 474011
|||||
717 GlySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIlePr 733
|||||
474012 GGATCCGCGCATGCGTGTGAGATCCGGAAGTGTGCGCATCGATTC 474061
|||||
733 oIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerL 750
|||||
474062 TATTGTATCCAAATGCTTCGAGACTGTTACGCTCTCTTCTTCTCTA 474111
|||||
750 yIleGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGly 766
|||||
474112 AACTGCAAGGATTTTCAGAACACAGACGCGTTTGTAGGAGAGTTCGGA 474161
|||||
767 GluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProI 783
|||||
474162 GAGATTCGGTCCCTTCTGCGAGCTCTTTTCAGAAATATTTCACTTCC 474211
|||||
783 eGlyIleThrPheGluLysLysSerGlnLysThrArgThrTyrTyrP 800
|||||
474212 AGGAATAACATTTGAAAAAATCCCAAAACACAGAACCTACTATTACT 474261
|||||
800 heLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyPro 816
|||||
474262 TTCTAGAGCGCTACATCCAGACCTGAACGCTGATGTGGAATCGGACCT 474311
|||||
817 ValValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLe 833
|||||
474312 GTAGTGTACTCAAAATGCGCTCTCTGGGATGCTCCTATGGCGAACTT 474361
|||||
833 uAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisA 850
|||||
474362 GGATTCAGAGCGCTACATGTTCCGGCTTACGAATCAAGAGCTCTACA 474411
|||||
850 rgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHis 866
|||||
474412 GACTTCAGAGCGCTTAAATGCTGCTTGTGCTGCTGGCGGCAAG.CAT. 474460
|||||
867 SerTyrSerLeuAspLeuGlyThr 874
|||||
474461 AGTTACTCCCTGGATCTGGGGACA 474484
|||||

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT.A64757

seq_documentation_block:

ID A64757 standard; DNA: 2601 BP.

XX AC A64757;

XX DT 02-FEB-2001 (first entry)

DE C. trachomatis pmpI gene coding sequence minus the signal sequence.

XX Chlamydial infection; sexually transmitted disease;

KW pelvic inflammatory disease; PID; tubal obstruction; infertility;

KW trachoma; blindness; acute respiratory tract infection;

KW atherosclerosis; coronary heart disease; antibacterial; ss.

XX Chlamydia trachomatis.

OS WO200034483-A2.

XX PN 15-JUN-2000.

XX PD

```

XX PF 08-DEC-1999; 99W0-US29012.
XX PR 08-DEC-1998; 98U5-0208277.
XX PR 08-APR-1999; 99U5-0288594.
XX PR 01-OCT-1999; 99U5-0410568.
XX PR 22-OCT-1999; 99U5-0426571.
XX PA (CORI-) CORIXA CORP.
XX PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
XX WPI: 2000-431303/37.
XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection
XX PT comprises immunogenic portion of Chlamydia antigen, which comprises
XX PT amino acid sequence encoded by polynucleotide sequence
XX PS Claim 1; Pages 198-199; 256pp; English.
XX CC The present invention relates to new nucleic acid sequences and the
XX CC proteins encoded by the nucleic acid sequences. The encoded proteins
XX CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
XX CC proteins are useful for the serodiagnosis and treatment of Chlamydia
XX CC infection..Chlamydiae are intracellular bacterial pathogens that are
XX CC responsible for a wide variety of human infections. C. trachomatis
XX CC infection is one of the most common sexually transmitted diseases and can
XX CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
XX CC and infertility. Trachoma due to ocular infection with C. trachomatis is
XX CC the leading cause of preventable blindness worldwide. C. pneumonia is a
XX CC major cause of acute respiratory tract infections in humans and is also
XX CC thought to play a role in the pathogenesis of atherosclerosis and
XX CC coronary heart disease. The present sequence is a nucleic acid sequence
XX CC isolated in the present invention.
XX SQ Sequence 2601 BP; 696 A; 641 C; 490 G; 769 T; 5 other;

```

alignment_scores:

Quality: 4373.00 Length: 857
Ratio: 5.127 Gaps: 0
Percent Similarity: 99.533 Percent Identity: 98.950

alignment_block:

US-09-677-752-4 x A64757

Align seg 1/1 to: A64757 from: 1 to: 2601

```

22 leupheGlyGlnAspProLeuGlyGluThrAlaLeuLeuThrLysAsnPr 38
|||||
28 CTCTTTGGCCAGGATCCCTTAGGTGAACCCGCCCTCCTCACTAAAAATCC 77
|||||
38 oAsnHisValValCysThrPhePheGluAspCysThrMetGluSerLeuP 55
|||||
78 TAATCATGCTGCTGTGTACATTTTTTTGAGGACTGTACCATGGAGAGCTCT 127
|||||
55 heProAlaLeuCysAlaHisAlaSerGlnAspProLeuTyrValLeu 71
|||||
128 TTCCTGCTCTTTGTGCTCATGCATCAAGACGATCCTTTGTATGTACTT 177
|||||
72 GlyAsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLy 88
|||||
178 GGAAATTCCTACTCTTGGTTCGTATCTTAAACTCCATATCAGGACCCCAA 227
|||||
88 sGluAlaLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||
228 AGAGCTCTTTTTTAAAGAAAAGAGATCTTTCCATTCAAAACCTTCGCT 277
|||||
105 heLeuSerPheThrAspCysSerSerLysGluSerSerProSerIleIle 121
|||||
278 TCCCTTCCTTCACAGATTGCTCTTCCCAAGGAAGACTCTCCTCTCTATT 327
|||||
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||||

```

738 ySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheS 755
 2178 TTCGGAGCTGTTACGTCCTTCTCTATTCTCTAAATCTCAAGGATTTT 2227
 755 erGlyThrGlnAspGlyPheGluSerSerGlyGluIleArgSerPhe 771
 2228 CAGGAACACAGGAGCGTTTTCAGGAGAGTTCGGGAGAGATTCGGTCTTT 2277
 772 SerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheG 788
 2278 TCTCCAGCTCTTTCAGAAATATTCTCTATAGGAATTAACATTGA 2327
 788 uLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyrI 805
 2328 AAAAAATCCCAAAAAACACGAACTACTATTCTTAGGAGCCCTACA 2377
 805 leGlnAspLeuLysArgAspValGluSerGlyProValValLeuLys 821
 2378 TCCAAGACCTGAACGCTGTGTGAATCGGACCTGTAGTGTACTCAA 2427
 822 AsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTy 838
 2428 AATCCGCTCTCTGGGATGCTCTCTATGCGAACTTGGATTCCAGGCTA 2477
 838 rMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuGlnThrLeuL 855
 2478 CATCTCCGGCTTACGAATCAAGAGCTCTACACAGACTTTCAGACGCTGT 2527
 855 euAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAsp 871
 2528 TAAATGTCTTGTGTGCTGCTGCGTGAACCAACCAATAGTACTCTCCCTGGAT 2577
 872 LeuGlyThrThrTyrArgPhe 878
 2578 CTGGGACCACTTACAGGTTT 2598

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C81914

seq_documentation_block:

C81914 standard; DNA; 273254 BP.

C81914;

27-FEB-2001 (first entry)

Chlamydia pneumoniae genome DNA.

Genome; diagnosis; vaccine; ds.

Chlamydia pneumoniae.

WO200027994-A2.

18-MAY-2000.

12-NOV-1999; 99WO-US26923.

12-NOV-1998; 98US-0108279.

08-APR-1999; 99US-0128606.

(REGC) UNIV CALIFORNIA.

Stephens R, Mitchell W, Kalman S, Davis R;

WPI: 2000-376516/32.

Isolated nucleic acid for use in diagnostic and analytical methods

PT encodes genomic sequence of Chlamydia pneumoniae

PS Claim 2: Page 128-320; 320pp; English.

XX This invention describes a novel nucleic acid (N1) encoding a Chlamydia

CC pneumoniae protein (P1), given in the specification. The isolated nucleic
 CC acid is useful for diagnostic and analytical methods, such as, protein may
 CC hybridization-based assays or amplification-based assays. The protein may
 CC be used for diagnostic purposes, for their enzymatic or structural
 CC activity, or as a vaccine. The invention also describes (1) a probe
 CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
 CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
 CC cassette comprising N1 under the transcriptional regulation of a
 CC transcriptional initiation region functional in an expression host, and a
 CC transcriptional termination region; (4) a cell comprising an expression
 CC cassette of (3) as part of an extrachromosomal element or integrated into
 CC the genome of a host cell as a result of induction of the expression
 CC cassette into the host cell, and the cellular progeny of the host cell;
 CC (5) a method for producing a P1 comprising growing a cell of (4) where
 CC the protein is expressed and isolating the protein free of other
 CC proteins; (6) a purified polypeptide composition comprising at least 50
 CC weight % of P1; and (7) a monoclonal antibody binding specifically to the
 CC peptide of (6).
 XX
 SQ Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

alignment_scores:

Quality: 947.50 Length: 910

Ratio: 1.771 Gaps: 32

Percent Similarity: 58.791 Percent Identity: 31.648

alignment_block:

US-09-677-752-4 x C81914/rev ..

Align seg 1/1 to reverse of: C81914 from: 1 to: 273254

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG 101
 105585 ACAATGCAGATGGAATCTATCTATACTTAACAGGGATGCTCAATCAC 105536
 101 nAspPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
 105535 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCTGCTGCTTTAAAGAACTA 105486
 117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
 105485 CTGGGAATCTTCTTCCAGGCCACGGCTACCAATTCTCTACAAAT 105436
 133AsnGlySe 135
 105435 ATCATGCGGGAGCGAACTGTACCTTTACCAATACAGCTCAATAAGCT 105386
 135 rMetSerPhe.....CysArgAsnH 142
 105385 TCTCTCTTTTCAGGATTCCTCTATTGTCACATAACAAACACCAAGATG 105336
 142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
 105335 CTACCACAGGAACAGAGCCCATCAAGTCCACAGGAGCTGTGCTCTTCAG 105286
 159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG 175
 105285 TCGAACTAT...AGTTGCTACTTTTGGCCAAAACCTTTCTAATGACAATGG 105239
 175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
 105238 AGCGCCCTCCAGGCGAGCTCTATCAGTCTATCGCTAAACCCCAAC...C 105192
 192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
 105191 TAAGCTTTCACAAACAAACACGCAACGCAAGGGGGTGCCTCTATTCC 105142
 209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG 224
 105141 ACGGAGGGATTACAATTAAACAATACGTTAAACTCAGCATCATTTCTGTA 105092
 224 yAsnSerAlaThrAsn..... 229

```

761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103313 TCAACAACGTAATACTACCTGGTACGATCTTTCGATACGGGTGATTGA 103264

777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnly 793
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103263 ATTAACGCTCTGTGCTATTGGAATACCTTCGAGAGATTCTCGAGAAA 103214

793 sThrArgThrTyrTyrTyPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103213 CGAGCTGGCTGTACGAGCTACTGCTACTACGTTCGGGTGCTATC 103164

810 rgAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103163 GTAAGAATCTGACTCCACGACAGCTCTCTAATCAACAATACCTCGTGG 103114

827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103113 AAAACTACAGGAACGAACTCTCAACAGCAAGCTGGTATCGGAAGACGAG 103064

838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103063 GATCTTTATGCTCTCT.....CCNAATCTTG 103035

852 lnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103034 AGCTCAAGTAACCTATCTATGAAATTCGTGGATCTTCACGACGCTAC 102985

869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102984 AATGCAGATCTTGGAGGTAAAGTCCAGTTC 102955

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30852

seq_documentation_block:

ID A30852 standard; DNA: 2784 BP.

XX AC A30852;

XX DT 29-AUG-2000 (first entry)

XX DE Chlamydia antigen CPN100638 gene open reading frame.

XX KW Chlamydia antigen; diagnosis: infection; community acquired pneumonia;
 KW therapy: upper respiratory tract disease; bronchitis; sinusitis;
 KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
 KW ds.

XX OS Chlamydia pneumoniae.

XX FH Key Location/Qualifiers

XX CDS 1..2784

XX FT /*tag= a

XX FT /partial

XX FT /product= chlamydia antigen CPN100638

XX FT /note= "no stop codon given"

XX PN WO200032794-A2.

XX PD 08-JUN-2000.

XX PF 01-DEC-1999; 99WO-CA01147.

XX PR 01-DEC-1998; 98US-0110339.

XX PR 01-DEC-1998; 98US-0110340.

XX PR 01-DEC-1998; 98US-0110427.

XX PR 01-DEC-1998; 98US-0110428.

XX PR 01-DEC-1998; 98US-0110438.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Murdin AD, Oomen RP, Wang J;

XX DR

WPI: 2000-412339/35.
 P-PSDB; Y90239.

XX DR

Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma.

XX PT

Claim 2; Fig 5; 174pp; English.

XX PS

This sequence encodes a Chlamydia antigen of the invention, designated CPN100638. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis), asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.

XX SQ

Sequence 2784 BP; 811 A; 689 C; 537 G; 747 T; 0 other:

alignment_scores:

Quality: 944.50 Length: 910

Ratio: 1.769 Gaps: 32

Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:

US-09-677-752-4 x A30852

Align seg 1/1 to: A30852 from: 1 to: 2784

85 ThrAspProLysGluAlaLeuPheLysGlyLysGlyAspLeuSerIleG1 101

154 ACAATGCCAGATGCAACTATCTATAATCTAACAGGGGATGCTCTCAATCAC 203

101 nasnPheArgPhe....LeuSerPheThrAspCysSerSerLysGluSerS 117

204 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCCTGCTTTTAAAGAACTA 253

117 erProSerIleIleHisGlnLysAsnGly....GlnLeuSerLeuArgAsn 132

254 CTGGGAATCTTTCTTCCAAAGGCCACGGCTACCAATTTCTCCTACAAAT 303

133AsnGlySe 135

304 ATCGATCGGGGACGAACTGCTACCTTTACCAATACAGCTGCAAAATAAGCT 353

135 rMetSerPhe.....CysArgAsnH 142

354 TCTCTCTTTTCAGGATTCCTCTATTGTGCACCTAATACAAACCCAGCAATG 403

142 isAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158

404 CTACCACAGGAACAGGACCCATCAAGTCCACAGGAGCTTGTCTATTTCAG 453

159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175

454 TCGAACTAT...AGTTGCTACTTTGGCCAAAACCTTTTCTTAATGACAATGG 500

175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192

501 AGCGGCGCTCCAAAGGAGCTCTCTACAGTCTATCGCTAAACCCCAAC...C 547

PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX P-PSDB; Y90239.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma -
XX
XX Claim 2; Fig 5; 174pp; English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
XX CPN100638. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to down regulate
XX expression of the proteins and may be used to treat infections. The
XX nucleic acids may also be used to produce the protein antigens they
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 2950 BP; 855 A; 724 C; 562 G; 809 T; 0 other;

alignment_scores:
Quality: 944.50 Length: 910
Ratio: 1.769 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x A30851

Align seg 1/1 to: A30851 from: 1 to: 2950

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG 101
|||
254 ACAATGCGAGATGGAAGTATCTATTAATCTAACAGGGGATGCTCAATCAC 303
|||
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||
304 CAATGCGAGATCTCCGACAGCTCTAACCCCTCTCTCTTAAAGAACTA 353
|||
117 erProSerIleLeHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
|||
354 CTGGGAATCTTTCTTCAAGGCCAGCGGTACCAATTTCTCTACAAAT 403
|||
133AsnGlySe 135
|||
404 ATCGATGCGGGAGCGAAGTGTACCTTTACCAATACAGCTGCCAATAGCT 453
|||
135 rMeLserPhe.....CysArgAsnH 142
|||
454 TCCTCTCTTTTTCAGGATTCCTCTATTGTGCTACTAATACAACCCAGCATG 503
|||
142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
|||
504 CTACCACAGGACAGGAGCCCAAGTCCACAGAGCTTGTCTTCTATTTCAG 553
|||

PR 01-DEC-1998; 98US-0110340.
PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
PR 01-DEC-1998; 98US-0110438.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX WPI: 2000-412339/35.
XX P-PSDB; Y90239.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma -
XX
XX Claim 2; Fig 5; 174pp; English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
XX CPN100638. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to down regulate
XX expression of the proteins and may be used to treat infections. The
XX nucleic acids may also be used to produce the protein antigens they
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 2950 BP; 855 A; 724 C; 562 G; 809 T; 0 other;

alignment_scores:
Quality: 944.50 Length: 910
Ratio: 1.769 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x A30851

Align seg 1/1 to: A30851 from: 1 to: 2950

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG 101
|||
254 ACAATGCGAGATGGAAGTATCTATTAATCTAACAGGGGATGCTCAATCAC 303
|||
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||
304 CAATGCGAGATCTCCGACAGCTCTAACCCCTCTCTCTTAAAGAACTA 353
|||
117 erProSerIleLeHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
|||
354 CTGGGAATCTTTCTTCAAGGCCAGCGGTACCAATTTCTCTACAAAT 403
|||
133AsnGlySe 135
|||
404 ATCGATGCGGGAGCGAAGTGTACCTTTACCAATACAGCTGCCAATAGCT 453
|||
135 rMeLserPhe.....CysArgAsnH 142
|||
454 TCCTCTCTTTTTCAGGATTCCTCTATTGTGCTACTAATACAACCCAGCATG 503
|||
142 isAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
|||
504 CTACCACAGGACAGGAGCCCAAGTCCACAGAGCTTGTCTTCTATTTCAG 553
|||

```
2279 TACCTTCTGGATCGAAGTGACGACGCTGCTCTCTTTGATGCTCAGAT 2328
      ::::::::::::::
695 acyTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPheS 712
      ||| :::::::::::::: |||
2329 CAGCTATCTATAGTAAAACTACTGAAACCTATTACACCCCAAGCAC 2378
      ||| :::::::::::::: |||
712 erLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
      ||| :::::::::::::: |||
2379 CAAAG...GGAGAGAGCTCGTGGTATATGACGCTCGCTCGTGAACCT 2425
      ::::::::::::::
729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      ::::::::::::::
2426 CGAGAGCTCCCTACCAACACACTGCTTTTAAAGCATGAGGCTCTCTCCACGC 2475
      ::::::::::::::
744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
      ::::::::::::::
2476 GTATTTCTCTTTCATCAAGTAGAAGCTTCGTACATACACCAAGATAGCT 2525
      ::::::::::::::
761 heGluGlu...SerSerGlyGluLeuArgSerPheSerAlaSerSerPhe 776
      ||| :::::::::::::: |||
2526 TCAAGAAGACGTAATACTACCTTGGTACGATCTCTTCGATACGGGTGATT 2575
      ||| :::::::::::::: |||
777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLy 793
      ||| :::::::::::::: |||
2576 ATTACGCTCTGTGCTTACGAAGCTACTGTCTACCTGCGGATCTCGAGAA 2625
      ||| :::::::::::::: |||
793 sThrArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
      ||| :::::::::::::: |||
2626 CGAGGTCGCTTACGAAGCTACTGTCTACCTGCGGATCTCTATC 2675
      ||| :::::::::::::: |||
810 rGaspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
      ||| :::::::::::::: |||
2676 GTAAGAATCTCTGACTGCGACGACACTCTCTTAATCAACAATACCTCGTG 2725
      ::::::::::::::
827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      ::::::::::::::
2726 AAACTACAGGAGCAATCTCTCAAGACACGCTGTATCGGAAGCAGG 2775
      |||
838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
      |||
2776 GATCTTTTATGCTTCTCT.....CCAAATCTTG 2804
      ::::::::::::::
852 lnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
      ::::::::::::::
2805 AGGTCAACAAGTAACCTATCTATGAAATTCGTGGATCTTCACGACGTAC 2854
      ::::::::::::::
869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
      ::::::::::::::
2855 AATGAGATCTTGGAGGTAACTTCAGTTT 2884
      ::::::::::::::
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X06822

```
seq_documentation_block:
ID   X06822 standard; DNA; 2787 BP.
XX
XX
AC   X06822;
XX
XX
DT   26-APR-1999 (first entry)
XX
DE   Chlamydia pneumoniae surface exposed protein Omp10 DNA.
XX
KW   Omp10; outer membrane protein 10; surface exposed protein; antigen;
KW   infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX
OS   Chlamydia pneumoniae.
XX
PN   W09858953-A2.
XX
PD   30-DEC-1998.
XX
PF   19-JUN-1998; 98WO-DK00266.
XX
```

```
PR 23-JUN-1997; 97DK-0000744.
XX
XX (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
XX Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
XX WPI: 1999-105610/09.
XX P-PSDB; W88423.
XX
XX Species-specific test for identifying mammals infected with
XX Chlamydia pneumoniae - comprises detecting antibodies specific for
XX outer membrane proteins of C. pneumoniae or nucleic acids encoding
XX these proteins
XX
XX Claim 6; Page 59; 115pp; English.
XX
XX This DNA sequence codes for the novel 98.4 kDa surface exposed
XX protein Omp10 (see W88423) of the human respiratory pathogen
XX Chlamydia pneumoniae. By generating antibodies against C.
XX pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
XX was obtained which reacted with outer membrane proteins. The
XX antibody was used to identify the genes (see X06816-27) encoding
XX Omp4-Omp15 proteins (see W88417-28) in an expression library of
XX C. pneumoniae DNA. The genes are situated in 2 gene clusters:
XX Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
XX the other, and encode polypeptides of about 89.6-100.3 kDa and
XX about 56.1 kDa. The invention provides a new species specific test
XX for identifying mammals (including humans) infected with Chlamydia
XX pneumoniae. The test comprises detecting antibodies specific for
XX Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
XX membrane proteins, especially by PCR. The proteins are also used
XX in the diagnosis of C. pneumoniae infection in mammals. The
XX nucleic acids and proteins can also be used in the immunization of
XX mammals, the nucleic acids being particularly useful as DNA
XX vaccines for effecting in vivo expression of antigens. The
XX vaccines may also prevent atherosclerosis and bronchial asthma,
XX which are possibly associated with C. pneumoniae.
XX
XX Sequence 2787 BP; 815 A; 689 C; 535 G; 748 T; 0 other;
XX
XX
```

```
alignment_scores:
Quality: 940.50 Length: 910
Ratio: 1.761 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.538

alignment_block:
US-09-677-752-4 x X06822 ..

Align seg 1/1 to: X06822 from: 1 to: 2787
```

```
85 ThrAspProLysGluAlaLeuPheLysGlyAspLeuSerIleG1 101
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
154 ACAATGCAGATGGAACATATCTATACTACAGGGGATGCTCAATC 203
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSers 117
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
204 CAATGCAGGATCTCCGACAGCTTAACCGCTTCTCTGCTTTAAAGAACTA 253
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
254 CTGGAATCTTTCTTTCCAGGCCACGCTACCAATTTCTCTCAATAAT 303
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
133 .....AsnGlySe 135
|||
304 ATCGATGCGGGAGGCACTGTACCTTTTACCANTACAGCTGCNAATAAGCT 353
|||
135 rMetSerPhe.....CysArgAsnH 142
:::|:::|:::|
354 TCTCTCCTTTTCAGGATCTCTCTATTTGTCATAATACAAACACGAAATG 403
|||
```

2129 CTCTCCATCTCCAGCATCTAGCAGCTTGTCTCTCTCCAGCTTGTACGC 2178
 691IleSerAlaGlnAl 695
 2179 TACCTTCTCGATCTGAAGTGAAGCAGCCTCTCTCTTTGATGTCAGAT 2228
 695 aCysTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPheS 712
 2229 CAGCTATATCTATAGTAAATATCTATGAAACCTATTACACCCCAAGCAC 2278
 712 eRLysGluGlyPheGlySerTyrHisSerValAlaValSerGlyGluVal 728
 2279 CAAG...GGAGAGCTCGTGGTATATGACGGTTGCGCTCTGGAACCT 2325
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
 2326 GCGAGCTCCCTACCACACACTGCTTTAAGCCATGAGGCTCTCTCCAGC 2375
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
 2376 GTATTTCTCTTCAAGTAGAGCTTCGTACATACACACAGATAGCT 2425
 761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
 2426 TCAAGAACCTTAATACTACTTGGTACCATCTTTCGATAGCGGTGATTTA 2475
 777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLys 793
 2476 ATTAACGTCTCTGTGCTTATTGGAATTACCTTCGAGAGATTCTCGAGAA 2525
 793 sThrArgThrTyrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
 2526 CGAGCGTGGCTTACGAAGCTACTGTCTACCTGCTGCCGATGCTATC 2575
 810 rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTyr 826
 2576 GTAAGAATCTCTGACTGCAGCAGCTCTCTTAATCAACAATACCTCGTGG 2625
 827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
 2626 AAACTACAGGAACGAATCTCTCAAGACAAGCTGTGTCGGAAGCAGG 2675
 838TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
 2676 GATCTTTTATGCTTCTCT.....CCAAATCTTG 2704
 852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
 2705 AGGTCAAGTAACCTATCTATGGAATTCGTGGATCTTCACGCAGCTAC 2754
 869 SerLeuAspLeuGlyThrThrTyrArgPhe 878
 2755 AATGCAGATCTTGGAGGTAAGTTCAGTTC 2784

seq_name: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X91990

seq_documentation_block:

ID X91990 standard; DNA; 1230025 BP.

XX AC X91990;

DT 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.

XX Chlamydia pneumoniae.

OS WO9927105-A2.

PD 03-JUN-1999.
 XX 20-NOV-1998; 98WO-1B01890.
 XX 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX (GEST) GENSET.
 XX Griffais R;
 PI WPI; 1999-357842/30.
 DR Genome sequence of Chlamydia pneumoniae
 XX
 PT Claim 1: Page 291-611; 1912pp; English.
 XX
 PS The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins Y34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC Y34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

alignment_scores:

Quality: 930.00 Length: 910

Ratio: 1.745 Gaps: 33

Percent Similarity: 58.571 Percent Identity: 31.538

alignment_block:

US-09-677-752-4 x X91990

Align seg 1/1 to: X91990 from: 1 to: 1230025

85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101
 510813 ACAATGCAGATGGAACATCTATCTATAATCTACAGGGGATGCTCAATCAC 510862
 101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSers 117
 510863 CAATCCAGGATCTCCGACAGCTCTAACCGCTTCTGCTTTAAAGAACTA 510912
 117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgasn 132
 510913 CTGGGAATCTTTCTTTCCAAGGCCACGCTACCAATTTCTCTCAAAAT 510962
 133AsnGlySe 135
 510963 ATCGATGCGGGAGCAACTGTACCTTTTACCAATACAGCTGCAATAAGCT 511012
 135 rMetSerPhe.....CysArgAsnH 142
 511013 TCTCTCTTTTCAGGATCTCTCTATTGTCACATAACAAACCAAGATG 511062
 142 IsAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
 511063 CTACACAGGACACAGAGCCATCAAGTCCACAGAGCTTGTCTATTTCAG 511112
 159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175
 511113 TCGAACTAT...AGTTGCTACTTTGGCCAAAACACTTTTCTATGACAATGG 511159
 175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
 511160 AGGCGCCCTCCCAAGCAGCTCTATCATGCTATCATCGCTAAACCCCAAC...C 511206

```

512935 CAAG...GGAGAGAGCTCGTGGTATTAATCAACGCTCCGCTCTGGAACCTT 512981
729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      ::::::::::::::: :::::::::::::::
512982 GCGAGCTCCCTACCAACACACTGCTTTAAGCCATGAGGCTCTCTTCCAGC 513031
744 rPheSerIlePheSerIleGlyLeuGlyPheSerGlyThrGlnAspGlyP 761
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513032 GTATTTCTCTTTCATCAAGATAGAAAGCTTCGTACATACACAGATAGCT 513081
761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513082 TCAAGAAGCACTAATACCTACCTGTCAGATCTTTCGATAGCGGTGATTTA 513131
777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLeuLysSerGlnLys 793
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513132 ATTAACGCTCTCTGCTGCTATTGGAATTACCTTCGAGAGATCTCGAGAA 513181
793 sThrArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysA 810
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513182 CGAGCGTGGCTCTTACGAAGCTACTGTCTACGTGCGGATGCTCTATC 513231
810 rGaspValGluSerGlyProValIleValLeuLysAsnAlaValSerTrp 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513232 GTAAGATCTGACTGCACACAGCTCTCTTAATCAACAATACCTCTCTGG 513281
827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      ::::::::::::::: :::::::::::::::
513282 AAAACTACAGGAAGCAATCTCTCAAGACAAGCTGGTATCGGAAGCAGG 513331
838 .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuG 852
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513332 GATCTTTATGCTCTCTCT.....CCAAATCTTG 513360
852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyr 868
      ::::::::::::::: :::::::::::::::
513361 AGGTCACAAGTAACCTATCTATCGAAATTCGTGATCTTCACGCGACTAC 513410
869 SerLeuAspLeuGlyThrThrArgPhe 878
      ::::::::::::::: :::::::::::::::
513411 AATGCAGATCTTGGAGGTAAAGTTCAGCTTC 513440

```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A30853

seq_documentation_block:

ID A30853 standard; DNA; 3000 BP.

```

AC A30853;
DT 29-AUG-2000 (first entry)
DE Chlamydia antigen CPN100639 full length coding sequence.
KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia;
KW therapy; upper respiratory tract disease; bronchitis; sinusitis;
KW asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma;
KW ds.
XX Chlamydia pneumoniae.
OS
XX
XX Key Location/Qualifiers
XX CDS 101..2893
XX /tag= a
XX /product= Chlamydia antigen CPN100639
XX
XX W0200032794-A2.
XX
XX 08-JUN-2000.
XX
XX 01-DEC-1999; 99WO-CA01147.
XX
XX 01-DEC-1998; 98US-0110339.
XX
XX 01-DEC-1998; 98US-0110340.

```

```

PR 01-DEC-1998; 98US-0110427.
PR 01-DEC-1998; 98US-0110428.
XX 01-DEC-1998; 98US-0110438.
PA (CONN-) CONNAUGHT LAB LTD.
XX
XX Murdin AD, Oomen RP, Wang J;
XX
XX WPI: 2000-412339/35.
XX P-PSDB: Y90240.
XX
XX Nucleic acids encoding polypeptide antigens from Chlamydia useful for
XX preventing, diagnosing and treating diseases such as community acquired
XX pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset
XX asthma -
XX
XX Claim 2: Fig 7: 174pp: English.
XX
XX This sequence encodes a Chlamydia antigen of the invention, designated
XX CPN100639. The nucleic acids (and their complementary sequences) may be
XX used as diagnostic agents for detecting the presence of nucleic acids
XX encoding Chlamydia antigens in samples according to standard methods,
XX and therefore, for diagnosing Chlamydia infections. For example, they may
XX be used as primers and probes for diagnostic polymerase chain reaction
XX (PCR) assays. Antisense sequences may be used to down regulate
XX expression of the proteins and may be used to treat infections. The
XX nucleic acids may also be used to produce the protein antigens they
XX encode according to standard recombinant DNA methodologies. The
XX proteins may then be used as antigens for the production of antibodies
XX (i.e. as vaccines) for preventing infection by Chlamydia. The
XX antibodies may also be used as diagnostic reagents for detecting
XX infections. Chlamydia is a pathogen implicated in the development of
XX (for example) community acquired pneumonia, upper respiratory tract
XX disease (especially bronchitis and sinusitis, asthmatic bronchitis,
XX adult-onset asthma and acute exacerbations of asthma in adults.
XX
XX Sequence 3000 BP: 858 A; 726 C; 592 G; 824 T; 0 other:

```

alignment_scores:

Quality: 912.50 Length: 840
Ratio: 1.702 Gaps: 26
Percent Similarity: 63.810 Percent Identity: 30.833

alignment_block:

US-09-677-752-4 X A30853

Align seg 1/1 to: A30853 from: 1 to: 3000

```

98 LeuSerIleGlnAsnPheArgPheLeuSerPhe.....Th 109
||| :: :::::::::: ||| ||| |||
455 CTAATCTTCACAGGATTTCTTAACCTTTCTTCTTCTTCAGCTCTCGGAAC 504
109 rAspCysSerSerGlySerSerProSerProSerIleIleHisGlnLysAsnG 126
| :::::::::: ||| :::::::::: |||
505 TACAGTTGCTTCAGGAAAAAGTAGTACTTTAACT.....TCTGCAG 542
126 IyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg..... 140
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
543 GAGCCTTAATCTTACCGATAATGGAACGATCTCTTTAGCCAAACGTC 592
141 ...AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSe 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
593 TCCAATGAAGCTAATAACAATGGCGGAGCGATCACCACAAAAACTCTTTC 642
156 rLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysG 173
| :::::::::: ||| :::::::::: |||
643 TATTCTGGGAATACCTCTTCTTCTATAACCTTCACTAGTAATAAGCGCAAAA 692
173 IyAsnGlyGlyAlaIle...GlnAlaGlnThrPheSerLeuSerArgAsn 188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
693 AATTAGGTGGAGCGATCTATAGCTCTCGCGCTGCAAGTATTTTCAGGAAC 742

```

```

2471  CCGTCTTCCAGGAGATATTCCCCCTCTTAAAGTTCAGCGCATCTACAG 3520
756   yThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPheSerA 773
      :  |||:::|||||:::|||||  |||:::|||||:::
2521  CCGCACACAAACTTTTAAAGAGAGTGGCGTGAAGCCCGTCTTTTGATG 2570
773   laSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys 789
      :  |||:::|||||:::|||||  |||:::|||||:::
2571  ATGGAGACCTAGTGAACCTCTATCCCTCTCGGCATTCGGTTAGAAAAA 2620
790   LysSerGlnLysThrArgThrTyrTyrTyrPheLeuGlyAlaTyrIleG1 806
      :  |||:::|||||:::|||||  |||:::|||||:::
2621  ATCTCCGAAGATGAAAAAATAATTTCGAGATTCTCTACGCTACATGG 2670
806   nAspLeuLysArgAspValGluSerGlyProValValLeuLeuLysAsnA 823
      :  |||:::|||||:::|||||  |||:::|||||:::
2671  TGATGTGTATCTAAAAATCCCGCTCGCGTACTTCTTAATGGTCAGTG 2720
823   lAlaSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
      :  |||:::|||||:::|||||  |||:::|||||:::
2721  GAGCCTCTGGACTTCGCTATGTGTAAACCTCGCACACAGCCTTCTTA 2770
840   PheArgLeuThrAsnGlnArgAlaLeuIis...ArgLeuGlnThrLeuLe 855
      :  |||:::|||||:::|||||  |||:::|||||:::
2771  GCAAGTGTGGAAGCCATCTGACTCTCTCCCTCATGTAGAACTCTCTGG 2820
855   uAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAspL 872
      :  |||:::|||||:::|||||  |||:::|||||:::
2821  GGAAGCTGCTTATGAGCTTCGTGCTCGACACACATCTACATGTAGATT 2870
872   euGlyThrThrTyrArgphe 878
      :  |||  |||  |||  |||
2871  GTGGCTAAGATACTACTATTC 2890

```

seq_name: /SIDS2/qcdata/geneseq/geneseq/NA1999.DAT:X06823

seq_documentation_block:

ID X06823 standard; DNA; 2793 BP.

[illegible]

AC X06823;

2 XX

DT 26-APR-

XX
DI 20 APR 1954

Chlamydia

DE
YY
CITY AND COUNTY OF LOS ANGELES

XX
MAY 11 1961

KW
KW

yy
KW
Direct

XX 50 prime 192

Chlamydia

XX
1,000,000

PN W098589

XX 2

PD 30-DEC-

XX

PF 19-JUN-

XX

PR 23-JUN-

XX

PA (BIRK/)

XX This DNA sequence codes for the novel 97.6 kDa surface exposed
CC protein Omp11 (see W88424) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see X06816-27) encoding
CC Omp4-Omp15 proteins (see W88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12.11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp15 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.

Sequence 2793 BP: 803 A: 681 C: 567 G: 742 T: 0 other: XX

alignment_scores:

Quality: 906.50

Ratio: 1.691

Percent Similarity:	63.810	Percent Identity:	30.714
Ratio:	1.001	Cups:	2

alignment_block:

US-09-677-752-4 x X06823

Align seg 1/1 to: X06823 from: 1 to: 2793

98 LeuSerIleGlnAsnPhcArgPheLeuSerPhe.....Th 109
 355 CTAACATTTACAGAGATTTCTACCTTTCTTCATTGCAGCTCTCTGGAAC 404
 109 rAspCysSerSerLysGluSerSerProSerIleIleHisGlnLysAsnG 126
 405 TACAGTTGCTTCAGGAAAACTACTTTAAGT.....TCTCGAG 442
 126 GlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg..... 140
 443 GAGCCTTAATCTTACCGAATATGTAAGCAGATTCTCTTTAGCCAAAACGTC 492
 141 ...AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSe 156
 493 TCCAATGAAGCTAATAACAATGCGGAGCGCATCACCAAAAAAAGCTTTTC 542
 156 rLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysG 173
 543 TATTTCTGGCAATACCTCTTCTATAACCTTCACCTAGTAATACGCCAAAA 592
 173 LysnGlyGlyAlaIle...GlnAlaGlnThrPheSerLeuSerArgAsn 188
 593 AATTAGGTGGAGCGATCTATAGCTCTCGGCTGCAAGTATTTTCAGGAAC 642
 189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAl 205
 643 ACCGGCCAGTTAGTCTTTATGAATAATAAGAGAAACTGGGGGCGGGC 692
 205 aIleCysCysSerAsnLeuIleCysSerGlyAsnVal.....AsnP 219
 693 TCTG.....GGCTTTTGAAGCCAGCTCTCTCGATTACTCAAAATAGCT 733
 219 roLeuPheThrGlyAsnSerAlaThrAsn.....GlyGly 231
 734 CCCTTTCTCTCTGGAAACACTGCACAGATGCTGCAGGCAAGGCGGG 783
 232 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSe 248


```

2368 CACGGTCTTCCAGGATATTTCCCTCTTAAGATTCCAGGCAGTCTAC 2417
756 GlyThrGlnAspGlyPheGluSerSerGlyGluIleArgSerPheSe 772
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2418 AGCGCGCAACAACAACTTAAGAGAGTGGCGCTGAAGCCGCTCTTTGA 2467
772 rAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluL 789
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2468 TGATGGAGACCTAGTGAACCTGCTATCCCTGTCGGCATTCGGTTAGAA 2517
789 yLysSerGlnLysThrArgThrTyTyTyTyPheLeuGlyAlaTyIle 805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2518 AAATCTCCGAAGATGAAGAAATAATTTCAGATTCTCTAGCCTACATT 2567
806 GlnAspLeuLysArgAspValGluSerGlyProValValLeuLeuLysAs 822
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2568 GGTGATGCTGTATCGTAA.....AAATCCCGCTTCGCTACTTCTCTAA 2610
822 nAlaVal.....SerTrpAspAlaProMetAlaAsnLeuAspSerA 836
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
2611 TGTCTCAGTGGAGCCCTCTGGACTTCGCTTAGTAAAGCCTCGCAGCAG 2660
836 rGAlaTyMetPheArgLeuThrAsnGlnArgAlaLeuLeuHisArgLeu.G1 852
   :|||||:|||||:|||||:|||||:|||||:|||||:
2661 AAGCCTCTTTAGCAAGTGTGGAAGCCATCTGACTCTCTCCCTCATCTAG 2710
852 nThrLeuLeu..AsnValSerCysValLeuArgGlyGlnSerHisSerTy 868
   :|||||:|||||:|||||:|||||:|||||:|||||:
2711 AACTCTCTGGGAAGCTGCTATGAGCTTCGTGGCTCAGCACACATCTAC 2760
869 SerLeuAspLeuGlyThrThrTyTyArgPhe 878
   :|||||:|||||:|||||:|||||:|||||:|||||:
2761 AATCTAGATTGTGGGCTAAGATACTCATTC 2790

```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A27021

seq_documentation_block:

ID A27021 standard; DNA; 3000 BP.

XX AC A27021;

XX DT 11-AUG-2000 (first entry)

XX DE Chlamydia pneumoniae 98kd putative outer membrane protein gene.

XX KW Chlamydia; antigen; vaccine; infection; outer membrane protein; ds.

XX OS Chlamydia pneumoniae.

XX FH Key Location/Qualifiers

XX FT CDS 101..2887

XX FT /*tag= a

XX FT /product= "98kDa putative outer membrane protein"

XX FT WO200026237-A2.

XX PD 11-MAY-2000.

XX PF 29-OCT-1999; 99WO-GB03579.

XX PR 29-OCT-1998; 98US-0106070.

XX PR 01-MAR-1999; 99US-0122066.

XX PR 27-OCT-1999; 99US-0428122.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Murdin AD, Oomen RP, Dunn PL;

XX DR WPI: 2000-365569/31.

XX DR P-PSDB; Y94327.

XX PT Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
PT for vaccination and protection against Chlamydia infection

XX Claim 1; Fig 1; 93pp; English.

XX The present sequence is the 98kDa putative outer membrane protein gene
XX from Chlamydia pneumoniae. The genomic sequence was amplified using two
XX PCR primers. The 5' primer contains a NotI restriction site, a ribosome
XX binding site, an initiation codon and a sequence close to the 5' end of
XX the 98kDa putative outer membrane protein coding sequence. The 3' primer
XX contains the sequence encoding the C-terminal sequence of the putative
XX outer membrane protein and a BsrGI restriction site. The stop codon was
XX excluded and an additional nucleotide was inserted to obtain an in-frame
XX C-terminal fusion with the Histidine tag. The PCR product was cloned
XX into a eukaryotic expression vector (pCA-Myc-His) by restricting both
XX the vector and the PCR product with NotI and BamHI and performing a
XX ligation reaction. This expression vector was injected intramuscularly
XX and intranasally into mice, which were subsequently inoculated with
XX Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
XX were lower than those of the controls. Thus the 98kDa putative outer
XX membrane protein can be used as a vaccine to provide protection against
XX Chlamydia infections, especially Chlamydia pneumoniae infections.
XX The polypeptide may also be administered orally to treat Chlamydia
XX infection. The present sequence may also be used in the
XX construction of attenuated Chlamydia strains that can over-express the
XX gene or express it in a non-toxic form.

XX Sequence 3000 BP; 871 A; 617 C; 634 G; 878 T; 0 other;

alignment_scores:
Quality: 866.00 Length: 865
Ratio: 1.701 Gaps: 24
Percent Similarity: 58.844 Percent Identity: 30.405

alignment_block:

US-09-677-752-4 x A27021 ..

Align seg 1/1 to: A27021 from: 1 to: 3000

```

74 SerTyrcysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
452 ACGTTTATAGGTTTCTTCGCTATCTTTTATTCGCTCTCTCGAAGTTC 501
90 aLeuphelysGluLysGlyAspLeuSerIleGlnAsnPheArgPheLeus 107
   :|||:|||||:|||||:|||||:|||||:|||||:
502 GATAACTACCGCAAGAGCGGTAGC..... 529
107 erPheThrAspCysSerSerLysGluSerSerProSerIleIleHisGln 123
   :|||||:|||||:|||||:|||||:|||||:|||||:
530 .....TGCTCTAGC..... 538
124 LysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysAr 140
   :|||||:|||||:|||||:|||||:|||||:|||||:
539 .....GGTAGCTTTCAGTTCGACAAAATGTCAGTTTGCTCTTCAGCAA 582
140 qAsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerL 157
   :|||||:|||||:|||||:|||||:|||||:|||||:
583 AAACCTTTTCAACGGGATATGCGGTGCTATCACCGCAAAACCTCTTCAT 632
157 euGlnHisAsnTyrlLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
   :|||:|||||:|||||:|||||:|||||:|||||:
633 TAACAGGGACTACAATGTCAGCTCTGTTTCTGAAAATACCTCTCAAG 682
174 AsnGlyGlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnVa 189
   :|||||:|||||:|||||:|||||:|||||:|||||:
683 AAAGCGGAGCCATTCAGACTTCGATCCCTTACCATTACTGGAACCA 732
189 lSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaI 206
   :|||:|||||:|||||:|||||:|||||:|||||:
733 AGGGGAAGTCTCTTTTCTGACATACCTTCTTCGGATCTTCGGAGCTGCAA 782
206 leCysCysSerAsnLeuIleCysSerGlyAsnValAsnProIlePhePhe 222
   :|||:|||||:|||||:|||||:|||||:|||||:
783 TTTTACAGAGCCGCTGCTACTATTCTTAATGCTAAAGCTTTCCTTT 832

```


851 euClnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSer 867
2772 TGGAGTGTTCACGTCAGTTGATCGAGTTCGCGAGTCTCTCGTAGC 2821

868 TyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
2822 TATGCTATTCGATCTTGGAGGAAGATTCGGATT 2854

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06820

seq_documentation_block:

ID X06820 standard: DNA: 2787 BP.

XX X06820:

26-APR-1999 (first entry)

Chlamydia pneumoniae surface exposed protein Omp8 DNA.

KW outer membrane protein 8; surface exposed protein; antigen;
infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
KW

OS Chlamydia pneumoniae.

PN WO9858953-A2

AA 30-DEC-1998

XX
PF 19-JUN-1998: 98WO-DK00266

23-JUN-1997: 97DK-0000744

PA (BIRK/) BIRKELUND S.

PA (CHRI//) CHRISTIANSEN G.

PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;

DR WPI: 1999-105610/09.

DR P-PSDB; W88421.

Species-specific test for identifying mammals infected with *Chlamydia pneumoniae* - comprises detecting antibodies specific for outer membrane proteins of *C. pneumoniae* or nucleic acids encoding these proteins

PS Claim 6: page 52-53: 115pp: English.

This DNA sequence codes for the novel 90.0 kDa surface exposed protein omp8 (see W88421) of the human respiratory pathogen *Chlamydia pneumoniae*. By generating antibodies against *C. pneumoniae* outer membrane complex, a polyclonal antibody (PAB 150) was obtained which reacted with outer membrane proteins. The antibody was used to identify the genes (see X06816-27) encoding Omp4-Omp5 proteins (see W88417-28) in an expression library of *C. pneumoniae* DNA. The genes are situated in 2 gene clusters: Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in the other, and encode polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The invention provides a new species specific test for identifying mammals (including humans) infected with *Chlamydia pneumoniae*. The test comprises detecting antibodies specific for Omp4-Omp5 or detecting nucleic acid fragments encoding these outer membrane proteins, especially by PCR. The proteins are also used in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting *in vivo* expression of antigens. The vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

SQ Sequence 2787 BP; 811 A; 583 C; 598 G; 795 T; 0 other;

alignment_scores:

argument_cscs: Quality: 857.00 Length: 865

Quadrat: 007.00	Wagen: 005
Ratio: 1.684	Gaps: 24

Percent Similarity:	Percent Identity:
58.844	30.173

alignment_block:

US-09-677-752-4 x X06820

Align seq 1/1 to: X06820 from: 1 to: 2787

74 SerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
352 ACGTTTATAGGGTTTCTTCGCTATCTTTATTCGCTCTCCTGGAAGTTC 401
90 aLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArgPheLeuS 107
402 GATAACTACCGCAAGGAGCGCTTAGC..... 439
107 exPheThrAspCysSerLysGluSerSerProSerIleIleHisGln 123
430TGCTCTAGC..... 438
124 LysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysAr 140
439GGTAGCTGAAGTTTGACAAAATGTCAGTTTGCCTCTACCAA 482
140 gAsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerL 157
483 AAAC'TTTTCAACCGGATAATGGCGGTCTATCACCGCAAAACCTTTCAT 532
157 euGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
533 TAACAGGGGACTACAATGTCAGCTCTGTTCTTGAAAATACCTCTCAAAG 582
174 AsnGlyGlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnVa 189
583 AAAGCGGAGCCATTCAGACTTCGCGATGCCCTTACCATTACTCGAAACCA 632
189 lserProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaI 206
633 AGGGAAGTCCTTTTCTGACAATACTTCTCGGATTCTGGAGCTGCAA 682
206 leCysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePhe 222
683 TTTTACAGAGCCCTCGTGACTATTCTTAATAAGTCTAAAGTTTCCTT 732
223 ThrGlyAsnSerAlaThr.....AS 229
733 ATTGACAATAGGTCACAGGACGAGCTCCTCAACAACGGGGATATGTC 782
229 nGlyGlyAlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlyS 246
783 AGGAGGCTCTACTCTGCT.....TATAAACTAGTACAGATACTA 823
246 erLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAla 262
824 AGGTCAACCTCTACTGGAAATCAGATGTTACTCTTCAGCAACAATACATCG 873
263 LysGluLysGlyGlyAlaIleTyrAlaLysHisMetValIleuArgTyrAs 279
874 ACAACAGCGGGAGGAGCTATCTATGTGAAAAAGCTCGAACTGGCTTCCGG 923
279 nGlyProValSerPheIleAsnAsnSer.....AlaLysI 291
924 AGGACTTACCCTATTTCAGTAGAAAATAGTCTCAATGGAGGTACAGCTCCTA 973
291 leGlyGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAla 307
974 AAGTGGAGCCATAGCTATCGAAGATAGTGGGAATTGAGTTTATCCGCC 1023
308 GlyGluGlySerValIleuPheGlnAsnAsnGlnArgThrSerAspGI 324

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z61508

seq_documentation_block:

ID Z61508 standard: DNA; 3050 BP.

XX

AC Z61508;

XX

DT 19-JUN-2000 (first entry)

XX

DE DNA encoding the CPN100394 polypeptide.

XX

KW CPN100394; Chlamydia infection; immune response; vaccine; ss.

XX

OS Chlamydia pneumoniae.

XX

FH Key Location/Qualifiers

FT CDS 101..2938

FT /*tag= a

FT sig_peptide 101..226

FT /*tag= b

XX

XX WO200011183-A2.

XX

XX 02-MAR-2000.

XX

PF 18-AUG-1999; 99MO-IB01449.

XX

PR 20-AUG-1998; 98US-0097187.

XX

PR 20-AUG-1998; 98US-0097188.

XX

PR 20-AUG-1998; 98US-0097189.

XX

PR 20-AUG-1998; 98US-0097190.

XX

PR 20-AUG-1998; 98US-0097195.

XX

PR 20-AUG-1998; 98US-0097196.

XX

PR 20-AUG-1998; 98US-0097197.

XX

PR 27-AUG-1998; 98US-0097191.

XX

PR 17-AUG-1999; 99US-0376770.

XX

XX (CONN-) CONNAUGHT LAB LTD.

XX

PI Murdin AD, Oomen RP;

XX

XX WPI: 2000-224703/19.

XX

DR P-PSDB; Y69368.

XX

XX Novel antigens and corresponding DNA molecules that can be used to

PT prevent, treat and diagnose disease caused by Chlamydia infection in

PT mammals, especially humans.

XX

NS Claim 1: Fig 13A-F; 201pp: English.

XX

CC Z61502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides are present in the bacterial membrane structure, in the external vicinity of the membrane structure, in the inclusion membrane structure, in the external vicinity of the inclusion membrane structure, and in the cytoplasm of the infected cell. The polypeptides may be used to prevent, treat and detect the presence of Chlamydia infection and/or the presence of Chlamydia in a sample. The polypeptides may also be used to induce an immune response in a mammal. The vaccine vector comprising the polynucleotides is used to induce an immune response in a mammal. Antibodies directed against the polypeptides may also be used therapeutically to treat and/or prevent a Chlamydia infection.

XX

SQ Sequence 3050 BP; 897 A; 622 C; 695 G; 836 T; 0 other;

alignment_scores:

Quality: 851.00

Ratio: 1.567

Percent Similarity: 62.702

Length: 866

Gaps: 25

Percent Identity: 28.637

alignment_block:

US-09-677-752-4 x Z61508

Align seg 1/1 to: Z61508 from: 1 to: 3050

```

74 SerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 AGCTTCACAGATTTCCTTCTCGTGATCACAGAAATCTCCAATAATCGC 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 aLeuPheLysGluLysGlyAspLeu.....SerIleG 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
499 TGTTACTACAGAAAGTAGCTAGTCAGTTAGTGCAGTCCAAGTGC 548
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
101 InAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSerSer 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
549 AAGATATAACACTCTAGTTCTTACAAGCAATGCTCTCTCGAAGATGGT 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
118 ProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnG1 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
599 GCCGTGATT.....AAGGAAACTCTGCTTGATTCAGGAATCAAAAA 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 ySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleS 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
643 TAGTGCATTTTGGACAAATACATCTTCGAAAAAAGAGGGGCGATCT 692
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151 er...AlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
693 CCAGGACTCAAGGACTTACCATAGAGATAACTTAGGGACGCTAAAGTTC 742
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
743 AATGAAACAAAGCAGTGACCTCAGGAGCGCCTTAGATTTAGGAGCGCG 792
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 acLntThrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgA 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
793 GTCTACATTTCACGCAACCATGAG.....TTGATATTTTCACAAA 833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 snArgAlaAsp.....LeuAsnGlyGlyAlaIleCysCysSer... 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
834 ATAAGACTTCTGGGAATGCTGCAATGCGGAGGCATAAATGCTCAGGG 883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 AsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAsnSe 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
884 GACCTTACATTTACTGATAACACTTCTTTGTTACTTCAAGAAATAGCAC 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
226 rAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrSerG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
934 AATCAGGATGCTGGAGCTTTGTGT.....A 959
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
243 LuLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSer 259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
960 GCACAGGAACCATAGCATTACCGGTAGTATTCTCAATGTGATAGGA 1009
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 AsnSerAlaLysGluLysGlyGlyAlaIleTyrAlaLysHisMet...Va 275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1010 AATACTTCAGACAAAAAGGAGGAGCGGATTTCTGCAGCTTCTCTCAAGT 1059
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 lLeuArgTyrAsnGlyProValSerPheIleAsnAsn.....S 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1060 TTTGGGAGGCGAGGCGCGCTCTCTTTCTTAATAACCTAGTACTCATG 1109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 erAlaLysIlleGlyAlaIleAlaIleGlnSerGlyGlySerLeuSer 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1110 CCACCCCTCTAGAGGTGCCATTTTATCAACACAGGAGGATCCTTGCAG 1159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 lLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgTh 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1160 CTCITTCACCTCAAGGAGGGGATATCGTATTTCGAGGGGAATCAGGTCACTAC 1209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
321 rSerAspGlnGlyLeuVal.....ArgAsnAlaIleTyrLeuGluLysA 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1210 AACAGCTCCAATGCTACCACCTAAGAGAAATGTAATTCACCTCGAGACA 1259
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
336 spAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...IleLeuPhe 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

762 uGluSerSerGlyLeuIleArgSerPheSerAlaSerSerPheArgAsnI 779
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2637 CGAAGCTCTAGTATGCGCGTGTGTTTAGTATGGAGGCTGCTTAACC 2686
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
779 leSerLeuProIleGlyIleThrPheGluLysLysSerGlnLysThrArg 795
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2687 TCTCGATTCTGTGGTGGCGAAATCTCGCAGGGGATATCGGAGATTCC 2736
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
796 ThrTyrTyrPheLeuGlyAlaThrIleGlnAspLeuLysArgAspVa 812
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2737 TACACCTATGATCTCTCAGGATCTTTTGTTCGATGCTATCGTAACA 2786
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
812 lGluSerGlyProValValLeuLeuLysAsnAlaValSerTrpAspAlap 829
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2787 TCCCAATCTACAGGACTCTTGTGATGACCCAGACTCTTGGAATAATC 2836
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
829 roMetAlaAsnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGln 845
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2837 GCGGTGGCAATCTTTCAAGACAGCATTTTACTGAGGGGTACCAAC... 2883
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 ArgAlaLeuHisArgLeuGlnThrLeuLeuAsnValSerCys..... 859
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2884 .....AACTAGCTACAACTCCCAATTGTGAGCTCTT 2915
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 .....ValLeuArgGlyGlnSerHisSerTyrSerLeuA 871
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2916 CGGACATTACGCTATGGAATCTCGTGTGATCTTCAAGGAACATAATGAT 2965
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
871 sPLeuGlyThrThrTyrArgPhe 878
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2966 ATGTTGTACCAACTCCGATTC 2988
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /SDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A30849

seq_documentation_block:

ID A30849 standard; DNA; 2950 BP.

AC A30849;

XX

DT 29-AUG-2000 (first entry)

XX

DE Chlamydia antigen CPN100635 full length coding sequence.

XX

KW Chlamydia antigen; diagnosis; infection; community acquired pneumonia; therapy; upper respiratory tract disease; bronchitis; sinusitis; asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma; ds.

KW

XX

OS Chlamydia pneumoniae.

XX

FH Key Location/Qualifiers

FT CDS 101..2887

FT /tag= a

FT /product= Chlamydia antigen CPN100635

FT sig_peptide 101..229

FT /tag= b

FT mat_peptide 230..2884

FT /tag= c

FT /note= "mature CPN100635"

XX

PN WO200032794-A2.

XX

PD 08-JUN-2000.

XX

PF 01-DEC-1999; 99WO-CA01147.

XX

PR 01-DEC-1998; 98US-0110339.

PR

PR 01-DEC-1998; 98US-0110340.

PR

PR 01-DEC-1998; 98US-0110427.

PR

PR 01-DEC-1998; 98US-0110428.

PR

XX 01-DEC-1998; 98US-0110438.

PA (CONN-) CONNAUGHT LAB LTD.

XX
PI Murdin AD, Oomen RP, Wang J;
XX

DR WPI: 2000-412339/35.
DR P-PSDB; Y90237, Y90238.
XX

PT Nucleic acids encoding polypeptide antigens from Chlamydia useful for preventing, diagnosing and treating diseases such as community acquired pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset asthma -
PT
XX

PS Claim 2; Fig 3: 174pp; English.
XX

CC This sequence encodes a Chlamydia antigen of the invention, designated CPN100635. The nucleic acids (and their complementary sequences) may be used as diagnostic agents for detecting the presence of nucleic acids encoding Chlamydia antigens in samples according to standard methods, and therefore, for diagnosing Chlamydia infections. For example, they may be used as primers and probes for diagnostic polymerase chain reaction (PCR) assays. Antisense sequences may be used to down regulate expression of the proteins and may be used to treat infections. The nucleic acids may also be used to produce the protein antigens they encode according to standard recombinant DNA methodologies. The proteins may then be used as antigens for the production of antibodies (i.e. as vaccines) for preventing infection by Chlamydia. The antibodies may also be used as diagnostic reagents for detecting infections. Chlamydia is a pathogen implicated in the development of (for example) community acquired pneumonia, upper respiratory tract disease (especially bronchitis and sinusitis, asthmatic bronchitis, adult-onset asthma and acute exacerbations of asthma in adults.
XX

SQ Sequence 2950 BP; 836 A; 593 C; 652 G; 869 T; 0 other;

alignment_scores:

Quality: 847.00 Length: 944

Ratio: 1.486 Gaps: 32

Percent Similarity: 60.381 Percent Identity: 30.191

alignment_block:

US-09-677-752-4 x A30849

Align seg 1/1 to: A30849 from: 1 to: 2950

12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProLe 28

158 TGTTCCACTGTTTTTGTGCACTGCTGAAATATAGGCCCTCTGATAG 207

28 u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValValC 43

208 CTTTGCAGGAAGTACTAACACAGGCCTATCTCTCTAAAAATACGACTA 257

43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54

258 CTGGAATAGACTATCTCTCAGAGGAGATATACCTGCAAAACCTTGGG 307

55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68

308 GATTTCGGCAGCTTTAACGAAGGGTGTGTTTTCTGCACACTACGGAATCTTT 357

68 uTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisIleT 85

358 AAGCTTTGCCGTAAGGGGTACTCA.....CTTCTCTTTTAAATATTA 401

85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleGln 101

402 AGTCTAGTCTGAAGG.....GCAGCCCTTTCTGTATACA 436

102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116

437 ACTGATAAAATCTGCGCTACAGGATTTTCGAGTCTTACTTCTTAGC 486

116 rSerProSerIleIle.....HisGlnL 124


```

738 Gly...SerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnG1 753
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
2332 GGACGTCGGTCTTGGGTGATACCCACCCACCCATTTCTAAACCTAGAGAT 2381
: : : : : : : : : : : : : : : : : : : : : : : : : : :
753 yPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyGluIleArgS 770
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2382 GATCTATGCACATCAGANTACCTTTAGGANAACGGCAGCAAGGCCGTT 2431
: : : : : : : : : : : : : : : : : : : : : : : : : : :
770 erPheSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThr 786
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2432 CTTTCCAAAGTTCAGACCTCTTCAATCTAGCGTCTCTGTAGGATAAAA 2481
: : : : : : : : : : : : : : : : : : : : : : : : : : :
787 PheGluLysSerGlnLysThrArgThrTyTyTyTyPheLeuGlyAl 803
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2482 TTTGAGAAATTCGCCATAAGTCT...ACGTATGATCTCTCCATA...GC 2525
: : : : : : : : : : : : : : : : : : : : : : : : : : :
803 atyIleGlnAspLeuLysArgAspValGluSerGlyProValValLeuL 820
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2526 TTACGTTCCCGATGTGATTCGTATGATCCAGGCTGCACGACAACTTTA 2575
: : : : : : : : : : : : : : : : : : : : : : : : : : :
820 eulysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArg 836
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2576 TGGTTCTGGCGATTCTTGTCACATGTGTACAAAGCTTGTTCTAGACAA 2625
: : : : : : : : : : : : : : : : : : : : : : : : : : :
837 AlaTyMetPheArgLeuThrAsnGlnArgAlaLeu...HisArgLeuG1 852
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
2626 GCTCTTCTTGTACGTGCTGGAAATCATCATGCTTTCGTTTCAAACTTGA 2675
: : : : : : : : : : : : : : : : : : : : : : : : : : :
852 nThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyrs 869
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2676 AGTTTCAGTCAGTTGAAGTCAGTTCGAGGTTCTTCTCGTAGCTATG 2725
: : : : : : : : : : : : : : : : : : : : : : : : : : :
869 erLeuAspLeuGlyThrThrArgPhe 878
: : : : : : : : : : : : : : : : : : : : : : : : : : :
2726 CTATCGATCTTCGAGGAAGATTCGGATT 2754
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT:X06827

```

seq_documentation_block:
ID X06827 standard; DNA; 2838 BP.
XX
AC X06827;
XX
DT 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp15 DNA.
XX
KW Omp15; Outer membrane protein 15; surface exposed protein; antigen;
KW infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX.
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998; 98WO-DK00266.
XX
PR 23-JUN-1997; 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
PA (CHRI/) CHRISTIANSEN G.
XX
PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mygind P;
XX
DR WPI: 1999-105610/09.
DR P-PSDB; W88428.
XX
XX Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins

```

```

XX Claim 6; Page 70-71; 115pp; English.
XX
CC This DNA sequence codes for the novel surface exposed protein Omp15
CC (see W88428) of the human respiratory pathogen Chlamydia pneumoniae.
CC By generating antibodies against C. pneumoniae outer membrane
CC complex, a polyclonal antibody (PAB 150) was obtained which reacted
CC with outer membrane proteins. The antibody was used to identify
CC the genes (see X06816-27) encoding Omp4-Omp15 proteins (see
CC W88417-28) in an expression library of Chlamydia pneumoniae DNA.
CC The genes are situated in 2 gene clusters: Omp12,11,10,5,4,13 and
CC 14 in one cluster and Omp6,7,8,9 and 15 in the other, and encode
CC polypeptides of about 89.6-100.3 kDa and about 56.1 kDa. The
CC invention provides a new species specific test for identifying
CC mammals (including humans) infected with Chlamydia pneumoniae. The
CC test comprises detecting antibodies specific for Omp4-Omp15 or
CC detecting nucleic acid fragments encoding these outer membrane
CC proteins, especially by PCR. The proteins are also used in the
CC diagnosis of C. pneumoniae infection in mammals. The nucleic acids
CC and proteins can also be used in the immunization of mammals, the
CC nucleic acids being particularly useful as DNA vaccines for
CC effecting in vivo expression of antigens. The vaccines may also
CC prevent atherosclerosis and bronchial asthma, which are possibly
CC associated with C. pneumoniae.
XX
SQ Sequence 2838 BP; 837 A; 591 C; 648 G; 762 T; 0 other:

```

```

alignment_scores:
Quality: 825.00 Length: 866
Ratio: 1.528 Gaps: 25
Percent Similarity: 62.356 Percent Identity: 28.176
alignment_block:
US-09-677-752-4 x X06827

```

```

Align seg 1/1 to: X06827 from: 1 to: 2838
74 SerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
349 AGCTTCACAGATTCTCTCTGCTGATCAGACAAATCTCCAAAATCCGC 398
: : : : : : : : : : : : : : : : : : : : : : : : : : :
90 aLeuPheLysGluLysGlyAspLeu.....SerIleG 101
: : : : : : : : : : : : : : : : : : : : : : : : : : :
399 TCTTACAGGAAAGTAGCTAGCTCAGTTTCAGGTGCAGTCCAACTGC 448
: : : : : : : : : : : : : : : : : : : : : : : : : : :
101 InAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSerSer 117
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
449 AGATATAAACACTCTAGTTCTTACAGCAATGCCTCTCTCGAAGATGGT 498
: : : : : : : : : : : : : : : : : : : : : : : : : : :
118 ProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnG1 134
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
499 GCGCTGATT.....AAGGAAACTCCTGCTTGATTCAGGGAATCAAAA 542
: : : : : : : : : : : : : : : : : : : : : : : : : : :
134 ySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleS 151
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
543 TAGTGGCAATTTTGGACAAAATACATCTTCGAAAAAAGGAGGGCGCATCT 592
: : : : : : : : : : : : : : : : : : : : : : : : : : :
151 er...AlaaspAlaPheSerLeuGlnHisAsnTyTyLeuPheThrAlaPhe 166
|| : : : : : : : : : : : : : : : : : : : : : : : : : :
593 CCACGACTCAAGGACTCACCATAGAGAAATAACTTAGGGACGCTAAAGTTC 642
: : : : : : : : : : : : : : : : : : : : : : : : : : :
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
: : : : : : : : : : : : : : : : : : : : : : : : : : :
643 ATGAAAAACAAGCACTGACCTCAGGAGGCGCTTAGATTTCAGACCCGC 692
: : : : : : : : : : : : : : : : : : : : : : : : : : :
180 aGlnThrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgA 197
: : : : : : : : : : : : : : : : : : : : : : : : : : :
693 GTCTACATTCACCTGCCGCAACCATGAG.....TTGATATTTTCACAA 733
: : : : : : : : : : : : : : : : : : : : : : : : : : :
197 snArgAlaasp.....LeuAsnGlyGlyAlaIleCysCysSer... 209
: : : : : : : : : : : : : : : : : : : : : : : : : : :
734 ATAAGACTTCTGGGAATGCTGCAATATGGCGGAGCCATAAATTCCTCAGGC 783
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```


681GlnAlaG 683
2118 GCACCTCATACCTACAAACCCAAAACATATTTGAGATTAGGAAGACAA 2167
683 lnLeuAlaThrSerLeuMetLys 690
2168 AGTTTCTGAACTAGCTATAGAAAAATTCCTAGCGAAATTCCTAGCC 2217
691 IleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGluLeuLysThrLy 707
2218 TTGCATGTCGAAGTTTCGTTCAGCCATTTCAGACAACCGTATGGAAACGCA 2267
707 sTyrArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAlav 724
2268 CTATACCTCATTTG...CCAGAATCCGAAGGTTCTTGGAGCAACGAGTGTA 2314
724 alSerGlyGluValCysAlaSerIlePro...IleValSerAsnGlySer 739
2315 TAGCTGGTGGTACGCTAGACCTTCCTTTGTTCTTTCCAAACCCACAT 2364
740 GlyLeuPheSerPheSerIlePheSerLysLeuGlnGlyPheSerGl 756
2365 CCTCTTTTCAGACCTTCATTCCACAGATGAAGTCGAAATGGTTATGT 2414
756 yThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPheSera 773
2415 ATCACAATAAGCTTCTTCGAAGCTCTAGTGATGGCGGTTTGTAGTA 2464
773 laSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys 789
2465 TTGGAAGGCTGCTTAACCTTCGATTCCTGTGGTGCGAAATTCCTGCAG 2514
790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyrIleGl 806
2515 GGGATATCGGAGATTCCTACACCTATCATCTCAGGATCTTTGTTC 2564
806 nAspLeuLysArgAspValGluSerGlyProvalValLeuLeuLysAsnA 823
2565 CGATGCTATCGTAACAATCCCAATCTACAGCGACTCTGTGATGAGCC 2614
823 laValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
2615 CAGACTCTTGGAAATTCGGTGCCAAATCTTCAAGACAGCATTTTTA 2664
840 PheArgLeuThrAsnGlnArgAlaLeuHisArgLeuGlnThrLeuLeuAs 856
2665 CTGAGGGGTAGCAAC.....AACTAGCTCTACAA 2693
856 nValSerCys.....ValLeuArgGlyGlnS 865
2694 CTCCAATTGTGAGCTCTTCGGACATTAGCTATGGAACCTCCGTGATCTT 2743
865 erHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
2744 CAAGGAACCTACAAATGTAGATGTGGTACCACAACTCCGATTC 2784

OM of: US-09-677-752-4 to: GenEmbl.* out_format : pfs
 Date: May 6, 2001 8:33 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame*_p2n.model -DEV=xlp
 -Q/cgn2/_USPTO.spool/US09677752/runat_04052001_132443_4919/app_query.fasta_1.948
 -DB-GenEmbl -OFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosom62 -TRANS-human4.0.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=25 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=200000000 -USER=US09677752_@CGN1_1_6443 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-677-752-4
 Query length: 878
 Database: GenEmbl.*
 Database sequences: 1283235
 Database length: 1216004940
 Search time (sec): 2529.030000

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_bal:AE001361	-	4509.00	5030.18	5.7e-272	5601	! AE001361 Chlamydia trachomatis
gb_bal:AE002294	+	3343.50	3941.98	2.3e-211	10908	! AE002294 Chlamydia muridarum
gb_bal:AE001628	+	947.50	1035.13	1.9e-49	10757	! AE001628 Chlamydia pneumoniae
gb_bal:AE002192	+	947.50	1033.40	2.4e-49	12676	! AE002192 Chlamydia pneumoniae
gb_bal:CPN133034	+	947.50	1030.13	3.6e-49	17280	! AJ133034 Chlamydia pneumoniae
gb_bal:AE002546	+	947.50	1000.01	1.7e-47	300650	! AP002546 Chlamydia pneumoniae
gb_bal:AE001839	+	940.50	1041.54	8.4e-50	2787	! A1839 Sequence 13 from Patent
gb_bal:AE002193	+	919.50	1004.52	9.7e-48	10026	! AE002193 Chlamydia pneumoniae
gb_bal:AE001627	+	915.50	995.75	3.0e-47	15068	! AE001627 Chlamydia pneumoniae
gb_bal:AE001841	+	906.50	1003.44	1.1e-47	2793	! A1841 Sequence 15 from Patent
gb_bal:CPU72499	+	896.50	986.26	1.0e-46	4926	! U72499 Chlamydia abortus
gb_bal:AE001829	+	883.00	977.05	3.3e-46	2815	! A1829 Sequence 3 from Patent
gb_bal:AE002192	+	883.00	961.18	2.5e-45	12676	! AE002192 Chlamydia pneumoniae
gb_bal:CPN133034	+	883.00	957.91	3.8e-45	17280	! AJ133034 Chlamydia pneumoniae
gb_bal:AE002546	+	883.00	927.78	1.8e-43	300650	! AP002546 Chlamydia pneumoniae
gb_bal:AE001628	+	866.00	943.87	2.3e-44	10757	! AE001628 Chlamydia pneumoniae
gb_bal:AE001835	+	857.00	948.04	1.4e-44	2787	! A1835 Sequence 9 from Patent
gb_bal:AE001851	+	853.50	943.34	2.5e-44	3000	! A1851 Sequence 25 from Patent
gb_bal:CPM954	+	853.50	935.98	6.4e-44	6030	! A1001311 Chlamydia pneumoniae
gb_bal:AE001587	+	852.00	923.71	3.1e-43	18448	! AE001587 Chlamydia pneumoniae
gb_bal:CPN133035	+	852.00	918.52	6.0e-43	25920	! AJ133035 Chlamydia pneumoniae
gb_bal:AE001837	+	849.50	938.18	4.8e-44	3200	! A1837 Sequence 1 from Patent
gb_bal:CPM954	+	849.50	931.50	1.1e-43	6030	! AJ001311 Chlamydia pneumoniae
gb_bal:AE002235	+	845.50	919.65	5.2e-43	12127	! AE002235 Chlamydia pneumoniae
gb_bal:AE002545	+	845.50	885.82	4.0e-41	299650	! AP002545 Chlamydia pneumoniae
gb_bal:AE001837	+	844.50	934.15	8.1e-44	2757	! A1837 Sequence 11 from Patent
gb_bal:AE001849	+	825.00	912.01	1.4e-42	2838	! A1849 Sequence 23 from Patent
gb_bal:AE001629	+	817.00	889.06	2.6e-41	10695	! AE001629 Chlamydia pneumoniae
gb_bal:AE002191	+	817.00	884.95	4.4e-41	15789	! AE002191 Chlamydia pneumoniae
gb_bal:CPM954	+	792.50	867.32	1.3e-40	6234	! U6943 Chlamydia abortus
gb_bal:CPM954	+	784.50	858.57	4.3e-39	6110	! U6942 Chlamydia abortus
gb_bal:AE002293	+	784.00	826.51	8.0e-38	14482	! AE002293 Chlamydia muridarum
gb_bal:AE001360	+	758.00	820.02	1.8e-37	14177	! AE001360 Chlamydia trachomatis
gb_bal:AE001585	+	740.00	802.77	1.7e-36	10766	! AE001585 Chlamydia pneumoniae
gb_bal:AE002237	+	739.00	802.38	1.8e-36	10044	! AE002237 Chlamydia pneumoniae
gb_bal:AE001831	+	735.00	810.46	6.2e-37	3052	! A1831 Sequence 5 from Patent
gb_bal:AE001586	+	708.50	767.68	1.5e-34	10574	! AE001586 Chlamydia pneumoniae
gb_bal:AE001833	+	705.50	779.42	3.3e-35	2526	! A1833 Sequence 7 from Patent
gb_bal:AE002189	+	575.50	619.09	2.9e-26	10336	! AE002189 Chlamydia pneumoniae
gb_bal:AE001631	+	574.50	616.22	4.1e-26	12083	! AE001631 Chlamydia pneumoniae
em_bal:AB033816	+	566.00	621.96	2.0e-26	2844	! AB033816 Chlamydia pneumoniae

gb_bal:AE001638 + 566.00 604.58 1.8e-25 14781 ! AE001638 Chlamydia pneumoniae
 gb_bal:AE002181 - 566.00 601.12 2.9e-25 20517 ! AE002181 Chlamydia pneumoniae
 gb_bal:AP002547 + 566.00 572.81 1.1e-23 300550 ! AP002547 Chlamydia pneumoniae
 gb_bal:AE002190 - 546.00 585.55 2.1e-24 10746 ! AE002190 Chlamydia pneumoniae

seq_name: gb_bal:AE001361

seq_documentation_block:
 LOCUS AE001361 5601 bp DNA BCT 30-OCT-2000
 DEFINITION Chlamydia trachomatis section 1 of 87 of the complete genome.

ACCESSION AE001361 AE001273
 VERSION AE001361.1 GI:3329348

KEYWORDS
 SOURCE Chlamydia trachomatis.

ORGANISM Chlamydia trachomatis
 Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

REFERENCE 1 (bases 1 to 5601)
 AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
 Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
 Koonin,E.V. and Davis,R.W.

TITLE Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis

JOURNAL Science 282 (5389), 754-759 (1998)
 MEDLINE 99000809
 PUBMED 9784136

REFERENCE 2 (bases 1 to 5601)
 AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C.J., Fan,J., Hyman,R.W.,
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.

TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
 MEDLINE 99206606
 PUBMED 10192388

REFERENCE 3 (bases 1 to 5601)
 AUTHORS Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R.,
 Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,
 Koonin,E.V. and Davis,R.W.

TITLE Direct Submission
 JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

FEATURES
 Location/Qualifiers
 1..5601
 /organism="Chlamydia trachomatis"
 /strain="D/UW-3/CX"
 /db_xref="taxon:813"
 323..640
 /gene="CT873"
 323..640
 /gene="CT873"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAC68471.1"
 /db_xref="GI:3329349"
 /translation="MSSNRMLATKKITVLLSKRKAGIRIDYCALADAVEYSVLWF
 LVGVAVYSSGVRAEILGETLSPGDRQAQDASLEASQVRLAGLMDYWDGDSRLK
 KKI"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

gene
 CDS
 Complement(815..3451)
 /gene="pmpl"
 /note="CT874"
 Complement(815..3451)
 /gene="pmpl"
 /codon_start=1
 /transl_table=11
 /product="Putative Outer Membrane Protein I"
 /protein_id="AAC68472.1"
 /db_xref="GI:3329350"

434 aLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMet 450
|||||
2151 TCTCTTCGGGCGCTCTCTCTCAGAGTCTCTCAGCTCTCTCTATTATG 2102
|||||
451 GluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrLe 467
|||||
2101 GAAGCGGAACTTCTTTAAAGTCTCTCTGATTTCGAAGTTAGCTACGCT 2052
|||||
467 uSerIleProLeuHisSerLeuAspThrGluLysSerValThrIleHisA 484
|||||
2051 AAGTATTCTCCCTTCATTCTCTAGATCTGAAAGGCGTAACATATCCACG 2002
|||||
484 laProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGlu 500
|||||
2001 CCCCTAACCTTTCTATCCCAAGATCTCTCTCTCTAATCTGGAGATGAG 1952
|||||
501 AsnPheTyrGluAsnValGluLeuSerLysGluGlnAsnIlePr 517
|||||
1951 AATTTTATCAAAATGTAGAGCTTCTCAGTAAAGAGCAAAACAATATCC 1902
|||||
517 oLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGlyA 534
|||||
1901 TCTCTTACTCTCTCTAAAGAGCAATCTCATTTACATCTTCTCTGATGGA 1852
|||||
534 snLeuSerSerHisPheGlyTyrGlnGlyAspThrPheSerTrpLys 550
|||||
1851 ACCTCTCTTCTCACTTTGGATATCAAGAGATGGACTTTTCTTGGAAA 1802
|||||
551 AspSerAspGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTy 567
|||||
1801 GATTCTGATCAAGGGCATCTCTGATTGCTAATTTGGACGCTAAAACTA 1752
|||||
567 rValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpA 584
|||||
1751 TGTGCTTCATCCAGAACGCTCAATCTACACTCGTTTGGAAACACTCTTTGGA 1702
|||||
584 snThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
|||||
1701 ACACCTATTCGGATATCAAGCTGTGAGTGTGATGATTAATACANTAGCG 1652
|||||
601 HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLe 617
|||||
1651 CACGAGGAGCCATCATCTATTGGAACTGGGATCTGCTGTTCTAATTT 1602
|||||
617 uPheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisA 634
|||||
1601 ATCTATGCTCACGACAGCTCTGGGAAACCTATCGATAATTGGCATCAT 1552
|||||
634 rgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHis 650
|||||
1551 GAAGCCTTGGCTACCTATTGGTATCAGTACTACAGTTTAGATGACCAT 1502
|||||
651 SerPheCysLeuAlaIaGlyGlnLeuLeuGlyLysSerSerAspSerPh 667
|||||
1501 TCTTCTGCTTGGCTGACGACAACTTACTCGGAAATCGTCCGATTCCTT 1452
|||||
667 eIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnL 684
|||||
1451 TATTACGCTCTACAGAAAGCACTCTATATAGCTACTCTACAGCGCAAC 1402
|||||
684 euAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSer 700
|||||
1401 TCGTACCCCTCTTAATGAAATCTCTGCACAGGATGCTTAATGAAAGT 1352
|||||
701 IleHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGI 717
|||||
1351 ATCCATGAGCTTAAACAAAATATCGCTCTCTCTAAAGAGGATTCGG 1302
|||||
717 ySerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProI 734
|||||
1301 ATCTGTCATAGCTGCTCAGTATCCGAGAGAGTGTGCGCATCGATTCCTA 1252
|||||

734 leValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerLys 750
|||||
1251 TTGTATCAATGTTCCGGACTGTTCCAGCTCTCTCTATTCTCTCTATAA 1202
|||||
751 LeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyGI 767
|||||
1201 CTCGAAGGATTTTCAGAACACAGGACGGTTTTCAGGAGAGTTCCGGAGA 1152
|||||
767 uIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleG 784
|||||
1151 GATTGGTCTCTTTCTGCCAGCTCTTTCCAGAAATATTTCACTTCTCTATGG 1102
|||||
784 lyIleThrPheGluLysLysSerGlnLysThrArgThrTyrTyrPhe 800
|||||
1101 GAATAACATTTGAAAAAATCCAAAAACACGAAACTACTATTACTTT 1052
|||||
801 LeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyProVa 817
|||||
1051 CTGGAGCCTTACATCCAAAGACCTAAAAACGTGTGTGGAATCGGACCTGT 1002
|||||
817 lValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuA 834
|||||
1001 AGTGTACTCAAAATGCCGTCTCTGGGATGCTCTCTATGGGACCTGG 952
|||||
834 spSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArg 850
|||||
951 ATTCTGGAGCCTACATGTTCCAGGCTTACGAATCAAGAGCTCTGCATAGA 902
|||||
851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSe 867
|||||
901 CTTTCAGACGCTGTTAAATGTCTCTTACGTACTGCGCGGCAAGCCATAG 852
|||||
867 rTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
|||||
851 TTACTCTCTGGATCTGGGACCACTTTACAGGTTT 818
|||||

seq_name: gb_ba1:AE002294

seq_documentation_block:

LOCUS AE002294 10908 bp DNA BCT 26-MAY-2000
DEFINITION Chlamydia muridarum, section 26 of 85 of the complete genome.
ACCESSION AE002294 AE002160
VERSION AE002294.1 GI:7190305

KEYWORDS

SOURCE Chlamydia muridarum.

ORGANISM

Chlamydia muridarum

REFERENCE

AUTHORS

Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

Genome sequences of Chlamydia trachomatis Morn and Chlamydia

pneumoniae AR39

Nucleic Acids Res. 28 (6), 1397-1406 (2000)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE

JOURNAL

FEATURES

source

1..10908

/organism="Chlamydia muridarum"

/db_xref="taxon:83360"

/note="synonym: Chlamydia trachomatis MoPn"

241..495

gene

```

CDS
complement(9731..10834)
/contig="rc0273"
/feature="conserved hypothetical protein; identified by
Glimmer2: putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="RAF39141.1"
/db_xref="GI:7190313"
/translation="MVKIMAPITPTSPQVKGSLRSRLTADPRHPKLRVYVDISLIAI
STLCIVSILITQSGGLAFAIPALAIICGLVTLIVSDIAESPKSKEVADPTAAVSL
PFILKTAGLAFMSAIVGGAVILANPLFLMGSLFALMSLHKVYQYLSNRQW
OKONIKIOESAANWKNLPKESKESLOTSRYSSLRKDKTRRNKPMGNKGSQVPA
SIANTERLSRFEVLHSSLLRQKELFNTNIIKELPNTKSLHTPLNRRSPSGSDS
DDVYTPRAGSSAETSLAGDLSIGSSSTSTSKTSTPKAKRRVYRSSERNARHRRN
KEDHPRNOEESDDSDSLPSPRRKKYRSRPK"
BASE COUNT      3210 a  1860 c  2587 g  3251 t
ORIGIN

alignment_scores:
  Quality: 3543.50      Length: 879
  Ratio: 4.429          Gaps: 2
  Percent Similarity: 91.013  Percent Identity: 77.816

alignment_block:
US-09-677-752-4 x AE002294/rev ..
Align seg 1/1 to reverse of: AE002294 from: 1 to: 10908

1 MetArgProAspHisMetAsnPheCysCysLeuCysAlaAlaIleLeuSe 17
3707 ATGCAGCAGATCATGTGAACCTTATGTTGCTATGCCCATCTTCTCTC 3658
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGluThrAlaLeuL 34
3657 CCCACACCAATCTCTTGTGAGGATCGCTTAGATAAATCCGCTCTTA 3608
34 euThrLysAsnProAsnHisValCysThrPhePheGluAspCysThr 50
3607 TTACTAAAAATCCTAACAGCATCGTCTGTACATTTTGGAAAGACTGCACC 3558
51 MetGluSerLeuPheProAlaLeuCysAlaHisAlaSerGlnAspAspPr 67
3557 ATGGGAAGACTTTCTCCCTGCTCTCTCTCATGCACAGCATGATGCC 3508
67 GluTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisI 84
3507 GCTATACATAATAGGAATACCAATACCTGTTGATATCCCAATCTCCACC 3458
84 leThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerile 100
3457 CCTCAACCAATCAAGAGCGGTTCTTAAAGAAAGAGGCGATCTTTCTATT 3408
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLys...GluSe 116
3407 CAAGACTTTTCGATTTCTTTCATTTACGGATTGCTCTCTCTACAGAAGA 3358
116 rSerProSerileleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnA 133
3357 CTCCTCTCTATCTCTACCAATAAATGGGAGCTATCTCTAAGAATA 3308
133 snGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAla 149
3307 ATGGAACATGAGCTTTTATCGCAATCATTCGCAAGGCTCTCGAGGAGCA 3258
150 IleSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPh 166
3257 CTCCTCCACAGAGCGATTGTTCTTACAAACAACAATCTATCTTTTCACAAAATTT 3208
166 eGluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrP 183
3207 TGAAGAAAATTTCTTCTGCAAAAATGGCGGTGCCATTCAGCTCAACATC 3158

```

```

183 heSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAla 199
3157 TTTCTTTGTCAAGAAATGCTCTCTCTCTTTTCCCGTAATCGCGCA 3108
200 AspLeuAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAs 216
3107 AATCTGAATGGAGGAGCTATTGTTGCCAAAATCTCATTTGTTTCAGAAA 3058
216 nValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyAlaI 233
3057 TGTAAACCTCTCTTTTCCCAATACTCTGCCCTAAATGGAGGAGCTA 3008
233 leCysCysIleSerLeuAsnThrSerGlyLysGlySerLeuSerLeu 249
3007 TTTGTTGTATTAATGAACAAATCTCTCAGAAAAGGATCGCTCTCTG 2958
250 AlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysG 266
2957 GCATATAACCAAGAGACTCTATTTCAGGCAATTCAGCAAGGAAAAGG 2908
266 yGlyAlaIleThrAlaLysHisMetValLeuArgTyrAsnGlyProVal 283
2907 GGGAGCCATTTACACAAACATATGTTTTCGCTCACAATGGACCCGTTT 2858
283 erPheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSer 299
2857 CTTTCGTAATAACAGTCTAAATAGGGGAGCTATTGCTATCCACTCA 2808
300 GlyGlySerLeuSerileLeuAlaGlyGluGlySerValLeuPheGlnAs 316
2807 GGAGGAAGTCTCTCCATCATTCAGGGGAGGCTCTGCTCTTTTCCAAA 2758
316 nasSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
2757 TAATCTCTGCTCTCTCTCATCAAGGTACGGTAAGAAATCGGATTACT 2708
333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAspIle 349
2707 TAGAGAAAATGCTCTGCTCTCTCTTTTGGAAAGTCCGCCATGGGATATC 2658
350 LeuPhePheAspProIleValGlnGluSerSerLysGluSerProLe 366
2657 CTTTCTTTTGATCCATTGTGCAAGAGTCTGTAAGTCTCTGAATCTCTAC 2608
366 uProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlas 383
2607 TACCTCAGCTCTA .....A 2594
383 erProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSer 399
2593 CGCCTTTACGATTCAGACAAACACACCGTCTGCTGTTATTTTCTAGT 2544
400 GluArgLeuSerGluGluLysThrProAspAsnLeuThrSerGlnLe 416
2543 GAAATCTATCCAAAGAGAGAAACCCAGCAATCTTATTCTTAAGAT 2494
416 uGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg 433
2493 ACAACAGCCAAATCGAGCTGCAATCTGGATGTTAGTCTTAAAGACCGAG 2444
433 laValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIle 449
2443 TCATTTCTTTCACCGCTCTTTTCCCAAGCCCCCAAGCTCTCTTAGTG 2394
450 MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaTh 466
2393 ATGGATGTAGGAGCTCTTTACACACATCTTCGGACTTCAATTAACAC 2344
466 rLeuSerIleProLeuHisSerLeuAspThrGlyLysSerValThrIleH 483
2343 TCTAAGCATTCCTCTCTTCTATAGATACAGAAAATAGTGTCTCTATCC 2294

```



```
seq_name: gb_bal:AE002192
seq_documentation_block:
LOCUS AE002192 12676 bp DNA BCT 30-MAY-2000
DEFINITION Chlamydia pneumoniae AR39, section 25 of 94 of the complete genome.
ACCESSION AE002192
VERSION AE002192.2
KEYWORDS Gi:8163405
SOURCE
ORGANISM Chlamydia pneumoniae AR39.
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 12676)
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 12676)
AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT On Jun 1, 2000 this sequence version replaced gi:7189226.
FEATURES
source
1..12676
/organism="Chlamydia pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(90..185)
/gene="CP0300"
complement(90..185)
/gene="CP0300"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38157.1"
/db_xref="GI:7189227"
/translation="MFTSLGRSAIFFTCSLKLHLRIGFELEIFQR"
complement(275..1819)
/gene="CP0301"
complement(275..1819)
/gene="CP0301"
/note="similar to GP:4376735; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38158.1"
/db_xref="GI:7189228"
/translation="MTILRNFLTCSALFLALPAAQVYVYVHSDGYNGAINNKSLEPK
ITCYPEGTSYIFLDVIRISNVKHDQACVFNRSNGLFPMGNRCNFTFHNLMTEGFG
AAISNRVGDITLILNFSYLAFTSAPLLPGOGGAYSLGSVMIESEVTFCCGNYSSW
SGAALTYPLLGSKASRSVNLUSGRNLYFRVDPQVGGAYSTINLITLRGSCFE
NNHAYHDVNSNGAIAIAPGSSISVSKSGDLIFKGNATSQDGNTHNSIHLSQSGAQF
KNLRVSESGYFYDPIHSHSESHKITDLVINAPKEGTEYEGTISFSGCLDDHEVCAE
NLITSTILODPTLAGGTLSIDCVTLQLHSFKQASSTLMSPGTTLICSGDARVONLH
ILIEDTDFVPIRAEDKDALVSLKLVKFAEAYVSVYDFPQKEAFTIPLELLGP
SFDLSLLGTTILRTQVTTENDAVRGFWNSLWEEPPPSLDKDRITPKTKVFLTNWP
EITSTP"
gene
CDS
complement(2077..4926)
/gene="CP0302"
complement(2077..4926)
/gene="CP0302"
/note="similar to GP:4376733; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38159.1"
/db_xref="GI:7189229"
/translation="MIYLCFCYIDANSLKNKSTIMKTSIPWLVSSVLAFLSCHLQSL
ANEALLPDDSFNGINDSTPTKTSATYSLTGDFVEYEPGKGTPLSDSCFKOTDN
LTFUNGHSILTFEIDAGTHAGAASTANKNLTFSGLSFSSSTVTTTGGQTL
SSAGVNLNIRKLVVAGNFSTADGGAIKGASFIILTGTSGDALFSNNSSTTKGAICAT
TAGARIANTGYVRFSLNASTSGAIDDECTSLSNKFLYFECNAKTTGGCAICNT
KASGPGLIISNNKTLIPASNVAETSGGAIHAKLALSSGGTFELRNNSVSSATPKGG
AISTDAGELSLSAETGNITVRNTLITTTGTDTPKRNAINIGSNGKFEELAAKNHT
IFFYDPTISGTSDDVLKINNGSAGALNPYQGTILFSGETLTADKLKADLNKSSFTQ
PVSLSGKLLQKGVLTLESTFSQEAAGSLMGDSGTLSTTAGSTITNLTGADVDSLV
LKQVSLTAKGASNKVIVSGKLNLDIEGNIYESHMFSDQLFLSKLITVADVDNTV
DISSLIPVPAEDPNSGEFGQGVNNTDTATNKEATWTKTGTFVPSERKSLV
CNTLWGFDTIRSIQQLVEIGATGMEHKQGFVSSMTNFKTGDNKRKRHRHSGGY
VIGSAITPKDGLTFEAFCHLFARDKDCFIHNNSTRTYGGTLFFKHSHLTQPNYLR
GRKFSESAIEKFEIPREIPLADVOVSFSDNRNMEHTYSLSESGSWSNCSIAAGIG
LDLPVLSNPHLPKTEIPOMKVMVYVSONSFESSDGRGFSIGRLNLNLSIPVGA
FVQDICSYTYDLISGFPVSDVYRNPOSTATLVMSPDMSKIRGNLSRQAFLLRGSN
NVTNSNCELFGYAMELRGSSRNYNVDVGTKLRF"
5026..7812
/gene="CP0303"
5026..7812
/gene="CP0303"
/note="similar to GP:4376729; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF38160.1"
/db_xref="GI:7189230"
/translation="MKSQFSLVLSSTLACFTSCSTVFAATAENTGPSDFGSDTNTG
TYTPKNTTGTIDYTLTGDTLQNLGDSAAATKGCDFSTESLSFAGKYSLSLNFKS
SAEGAAALSVTDKLNLSLTFEALAPSSVITTPSGKAVKCGDGLTFDNNGTILF
KODYCEANGGAIATKNLSKNGTSISGATKGGGAICATGVDITNNTAPT
LFSNIAEAGGAINSTGCTITGTSVFSNSVATAGNAGLSDGADVTISGNOS
VTFSGNAVANGGAIYAKKLTILASGGGGSFNSNIVOGTTAGNAGSILAAAGCSL
SAEAGDITFNGNAIVATTPTTKRNSIDIGSTAKITNLRAISCHSIFFDPTANTAA
DSTDTLNLNKADAGNSTDYSGSIVFSEKLSDEAKVADNLTSLKQPVTLTAGNLVL
KRGVTLDTKFTTAGSSVIMDAGTTLKASTEEVTLTGLSPVSLGKGVVIAASA
ASKNVALSGPILLDNOGNAYENHDLGKTQDFSVOLGALGTATTDVPAVPTVATPT
HYGYGTWGTWVDDTASTPKTKTATLAWNTGYLPNPEROGPLVNSLWGSFSDIOA
IQGVIERAALICSDRGFWAAGVAFNLDKDKGEXRKHGKGGVAGGAATCSENL
ISFACOLFGSDKDFLVAKNHDTYAGFYIOHTCSGFIGCLLDKLPKGSWKPLV
LEGOLAYSHVSNDLATKYTAYPEVKSGNNAFNMMLGASHSYPEYLHCFDTAPYI
KLNLYIRODSESEKGTGRSDFDNLNLSLPIGVKFEKFDNDLISYVDP
LIRNDPKCTTALVISGASWETVANNLARQALQVRAGSHVAFSPMEVLQGVFEVRGS
SLRYNVDLGGKQF"
complement(8170..8292)
/gene="CP0304"
complement(8170..8292)
/gene="CP0304"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38161.1"
/db_xref="GI:7189231"
/translation="MDLSRWLNKLKTRKQLQDNKIKYCKSYCLRGISQILGWV"
8329..9444
/gene="CP0305"
8329..9444
/gene="CP0305"
/note="conserved hypothetical protein; identified by
```

399 erGluArgLeuSerGluGluGluThr...ProAspAsnLeuThrSer 414
 11313 GAGAGAAGCTCTCGGAAGCAGAGCTGAGAGCTGATTAATCTCAAACTT 11264
 415 GlnLeuGlnProIleGluLeuLysSerGlyAlaGluLeuLysAs 431
 11263 ACAATTACAGCACTCTACTCTGCGGAGGCACTCTCTTAATC 11214
 431 pArGAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuL 448
 11213 AGGAGTCACTCTAGTTGCGCAAGTCTTTTCGCAATCTCGGGCTCTACCC 11164
 448 euLleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 454
 11163 TCCTCATGATGAGGAGGACACATTAGAACCCCTGATGGGATCACTATC 11114
 465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
 11113 AATAATCTTCTCAATGTAGATTCCTTAAGAGACACNAGAGGCTAC 11064
 481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer. 497
 11063 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTTATCTGGATCGC 11017
 498GlyAspGluAsnPheTyrgluAsnValGluLeuLeu 509
 11016 TCTCTCTGTAGATCTCTCTGGAATGTCTACGAAGATGTC..... 10976
 510 SerLysGluGlnAsn.....IleProLeuLeuThrLeuSerLy 523
 10975FCTTGGAAATACCTCAAGTCTTTCTGTCTCACTCTTACTGC 10932
 523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSer..... 537
 10931 TGACGACCCCGCAATATTCACATCAGAC.....TTAGCTGCTGATC 10888
 538HisPheGlyTyrgluGlnGlyAspThrPhe 547
 10887 CCCTAGAAAAAATCTCTATCTGCGGATACCAAGGAATGGGCATTA 10838
 548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
 10837 TCCTGGCAAGAGGATCTGCGACTAATATCCAAACGACGACTCTTACTGT 10788
 562 pThrProLysAsnTyrrValProHisProGluArgGlnSerThrLeuValA 579
 10787 GACAAACAGGATACAAATCCGAATCTGAGCGTGTGGAACCTTAGTTG 10738
 579 laAsnThrLeuTrpAsnThrTyrrSerAspMetGlnAlaValGlnSerMet 595
 10737 CTAACACATATGGGGATCTCTTTCTGATGTGCGCTCCATACACAGCTT 10688
 596 ileAsnThrIleAlaHisGlyGlyAlaTyrrLeuPheGlyThrTrpGlySe 612
 10687 GTACCCCTAAAGTACCCCAATCTCAGAAACTCGCGGCATCTGGTGTA 10638
 612 rAlaValSerAsnLeuPheTyrrAlaHisAspSerSerGlyLysProIleA 629
 10637 AGGATCTCGAATCTTC.....CATAAGATAGCAGCAAGATAAATA 10594
 629 sAsnTrpHisHisArgSerLeuGlyTyrrLeuPheGlyIleSerThrHis 645
 10593 AAGGTTTTCCGCACATAATGCGAGCTATGTGTAGGAGGCACTACACA 10544
 646 SerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLy 662
 10543 TTAGCTTCTGATATCTTATCACTGACGACCTTCGCCAATATTTCGGGA 10494
 662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyrrIleAla 679
 10493 AGATAGATACATCTTATAATAAATAATAGAGCTTCTCCCTATGACGCTT 10444

679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
 10443 CTCTCCATCTCCAGCATCTAGCAGCTTCTCTCTCAAGCTTGTTCACGC 10394
 691IleSerAlaGlnAl 695
 10393 TACCTTCTCTGGATCTGAAAGTGAGCAGCTGCTCTTTGATGCTCGCAT 10344
 695 acyTyrrAsnGluSerIleHisGluLeuLysThrLysTyrrArgSerPhe 712
 10343 CAGCTATATCTATAGTAAATACTATGAACCTATTACACCAAGCAC 10294
 712 erLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyLeuVal 728
 10293 CAAG...GGAGAGACTCTGTGTATTAATGACGTTGCGCTCTGGAACCT 10247
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
 10246 GCGAGCTCTCTACACACACTGCTTTAAGCATGAGGGTCTCTTCCACGC 10197
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
 10196 GTATTTTCTCTTTCATCAAGTAGAAGCTTCGTACATACACCAAGATAGCT 10147
 761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
 10146 TCAAGAACCTAATACTACTCTGTGACGATCTTTCGATACCGCTGATTTA 10097
 777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLy 793
 10096 ATTAAGCTCTCTGTGCTATTTGGAATTAACCTTCGAGAGATTCTCGAGAA 10047
 793 sThrArgThrTyrrTyrrPheLeuGlyAlaTyrrIleGlnAspLeuLysA 810
 10046 CGACGCTGCTCTTACGAAGCTACTGTCTACGTTCGCGATGCTCATC 9997
 810 rGaspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
 9996 GTAAGAATCTGACTCCACGACACTCTCTATCAACATACCTCTCTGG 9947
 827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
 9946 AAAACTACAGGACGAATCTCTCAAGACAAGCTGTATCGGAAGACGAG 9897
 838TyrMetPheArgLeuThrAsnGlnAlaLeuHisArgLeuG 852
 9896 GATCTTTTATGCTCTCT.....CCAAATCTTG 9868
 852 lnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTy 868
 9867 AGGTCAAGTAACCTATCTATGGAATTCGTGATCTTCACGACGCTAC 9818
 869 SerLeuAspLeuGlyThrThrTyrrArgPhe 878
 9817 AATGCAGATCTTGGAGTAAAGTTCAGTTC 9788

seq_name: gb_ba2:CPN133034

seq_documentation_block:

LOCUS CPN133034 17280 bp DNA BCT 14-NOV-2000
 Chlamydia pneumoniae GGAI-protein family cluster A, genes omp12
 DEFINITION (partial), omp11, omp10, omp5, omp4, omp3 and omp14 (partial) and
 ORF1, strain VR1310.

ACCESSION AJ133034
 VERSION AJ133034.1 GI:4455881
 KEYWORDS hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13
 gene; omp14 gene; omp4 gene; omp5 gene; ORF1; outer membrane
 protein; outer membrane protein 11; outer membrane protein 12;
 outer membrane protein 14; outer membrane protein 4; outer membrane
 protein 5.

SOURCE Chlamydia pneumoniae.
 ORGANISM Chlamydia pneumoniae
 Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

```
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/db_xref="SPTREMBL:O86164"
/translation="MKTSIPWLVSSVLAFLSCHLOSILANEELLSPDDSFNFGNIDSGTF
TPKTSATITGVDFYEPGKGTPLSDSCFKOTDNLITFLNGHSLTFPGFIDACTHA
GAAATSTANKNITFTSGFSLSDSPSTVTITGGTSSAGGVNLEIRKLVAGNES
TAGDAIKAGSLITGSGDALFNNSSSTKGAATATAGARIANNTGYVRFNLNIAF
TSGAIDDEGTSILNNKFLYEGNAKTTGCAICNTKAGSPGLIISNNKTLIFASN
VRYTSGGAHAKKALASGGGTEPLRNWVSATPKGKAISIDASGELSLSAETGNITF
VAENITGSGTTPARNAINIGSKGFTFELRAKNHTIFFDPTISEGTSDDVLKINN
GSAGALNPQGTIIIFSGTTLADKLKADNLKSPQVPSUSGKLLQKGVLESTG
FSQAGSLTGGTITLSTAGSITITNLGINVDSGLKQPSVSTAKGASKNVLSGK
LNLIDIEGNIYESHMFSDQLSKLITVDADVDNTDVISSLIPVAPEDPNSEYFGQ
QNNYMTTDTATNTKEATATWTKCFVPSPERKSALVNTLWGVFTDRLSLOQLVEIG
ATGMEHKGQFVSSWNTNLFHKTGDENRKGFRHTSGGYVIGSANTPKDDLTTFACHL
FARDKDCFIANNSTYGGTILFFKHSHTLQPNYLRLGRKAFSESAIEKFPRIPLAL
DVQVFSHSDNRMETHTYSLSESGSNQECIAGIGLIDLPLFVLSNPHPIFKTLPOM
KVMYVVSQNSGFFSSDGRGFSIGRLNLNLSIPVGAKEVQGDIGDSYTYDLSGFEVSD
VIRNPQSTATLVMSPDMSKIRGGNLSRQAFLLRGNNVYVNSNCELPCHYAMELRGS
SRNYNDVGTKLRF"
14579..14644
/gene="omp13"
14579..16123
/gene="omp13"
14579..16123
/gene="omp13"
/codon_start=1
/transl_table=11
/protein_id="CAB37073.1"
/db_xref="GI:4455888"
/db_xref="SPTREMBL:O923D6"
/translation="MTILRNELTCSALFLALPAARQVYVLLHESDGYNGAINKKSLEPK
ITCYPEGSTYFLDDVRISNVKHDQDAGVFINRSGNLFNFMGNRCNFTFHNLMTEGFG
AAISNRGDTITLNSFYLAFTSAPLQPGQGAISYLSGVNMIENSEEFTGNYSSW
SNGAAYTYPLLSKASRPVNLGRYLVFRDQVSGQGAISTHNLTLTTRGPSQCFE
NHHAYHDVNSGGAIAIAPGGSISLSVKSGDILFKGNTASQDCNTIHNSIHLQSCAF
KNLRAVSEGVYFPDIPSHSESKITDLVINAECKEYEGTISFSGLCDDHEVCAL
NLTSLILOYDTLAGTISLSGVTIQLHSFKQEAESTLMTSPGTTLLSGDARVONLI
ILITEDTNFVPRIRAEKDALVLEKLVAFYWSYDFPFQFREAITIPULELLG
```

sig_peptide

gene

CDS

alignment_scores:

```
Quality: 947.50 Length: 910
Ratio: 1.771 Gaps: 32
Percent Similarity: 58.791 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x CPN133034 ..

Align seg 1/1 to: CPN133034 from: 1 to: 17280
```

```
85 ThrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleG1 101
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
3980 ACAATGCAGATGCAATATCTAATACTTAACAGGGGATGCTCAATCAC 4029
101 nAsnPheArgPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
4030 CAATGCAGGATCTCCGACAGCTCTAACCGCTTCTGCTTTAAGAAACTA 4079
117 erProSerIleIleHisGlnLysAsnGly...GlnLeuSerLeuArgAsn 132
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
4080 CTGGGAATCTTTCTTCCAGCGCCAGCGCTACCAATTTCTCTACAAAT. 4129
133 ..... : : : : : : : : : : : : : : : : : : : : : : : : :
4130 ATCGATCGCGGAGCAACTGTACCTTTACCAATACAGCTGCAAAATAGCT 4179
135 rMetSerPhe..... : : : : : : : : : : : : : : : : : : : : :
4180 TCTCTCTCTTTTCAGGATTCCTATTTGTCATAATAACACACGGAATG 4229
142 isAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 158
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : |||
```

```
4230 CTACCACAGCAACAGGAGCCATCAAGTCCACAGGAGCTTGTCTTATTCAG 4279
159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnG1 175
||||| ||| : : : : : ||| : : : : : ||| : : : : :
4280 TCGAATAT...AGTTGCTACTTTGGCCAAACTTTTCTAATGACAAATGG 4326
175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerProI 192
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4327 AGCGCGCCTCAAGGAGCTCTATCAGTCTATCGCTAAACCCCAAC...C 4373
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
4374 TAACGTTTTCGCAAAACAAACAAACGCAACGAAAGGGGTGCCCTCTATTC 4423
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
: : : : : ||| : : : : : : : : : : : : : : : : : : : : :
4424 ACGGAGGGATTAACAATTACAAATACGTTTAACTCAGCATCATTTCTGA 4473
224 yAsnSerAlaThrAsn..... : : : : : : : : : : : : : : :
4474 AATACCGCGCGGCAACAATGGCGGAGCCATTTACACGGAAGCTAGCAGTT 4523
229 ..... : : : : : : : : : : : : : : : : : : : : :
4524 TTATTAGCAGCAACAAGCAATTAGCTTTTATAACAATAGTGTGACCGCA 4573
230 .....GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
4574 ACCTCAGCTACAGGGGAGCCATTTACTGTAGTAGT.....ACATCAGC 4617
242 rGluLysGlySerLeuAlaCysAsnGlnGluThrLeuPheAlas 259
: ||| : : : : : : : : : : : ||| : : : : :
4618 CCCCACCAACCTCTTAATCTATCAGACACGGGGAACCTGAACCTTTATAG 4667
259 erAsnSerAlaLysGlyGlyAlaIleTyrAlaLysHisMetVal 275
: : : : : : : : : : : : : : : : : : : : : : : : : : :
4668 GAAATACAGCAATTAATACGTGTGGTGGCGGATTTTACTGACAATCTAGTT 4717
276 LeuArgTyrAsnGlyProValSerPheIleAsnAsnSer..... 288
||| : : : : : : : : : : : ||| : : : : :
4718 CTTTCTCTCGAGGACCTACGCTTTTAAACCAACTCTGCTATAGATAC 4767
289 ....AlaLysIleGlyGlyAlaIleAlaIleGlnSerGlyGlySerLeus 304
||| : : : : : : : : : : : ||| : : : : :
4768 TGCAGCTCCCTTAGAGGAGCAATTCGCTGCTGACTCTGGATCTTGA 4817
304 erIleLeuAlaGlyGlySerValLeuPheGlnAsnAsnSer..... 318
: : : : : ||| : : : : : : : : : : : : : : : : : :
4818 GTCCTTCGGCTCTTGGTGGAGACATCACTTTTGAAGGAAACACAGTAGTC 4867
319 ...GlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu... 333
4868 AAGGAGCTCTTCGAGTCAGACCACTACCAGAAATCTTATAACATCGG 4917
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...I 349
: : : : : ||| : : : : : : : : : : : ||| : : : : :
4918 AAACACCAATGCTAAGATTGTACAGCTCCGAGCCTCTCAAGGCAATACTA 4967
349 leLeuPhePheAspProIleValGlnGluSerSerSerLysGluSerPro 365
||| : : : : : : : : : : : ||| : : : : :
4968 TCTACTTCTATGATCCTATAACA..... : : : : : : : : : : :
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
4991 .....ACTAGCATCACTGCAGCTCTCTCAGATGCTCTAAACTTAAATGG 5034
382 aserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerS 399
: : : : : ||| : : : : : : : : : : : ||| : : : : :
5035 TCCTGACCTTCAGGGAATCTCGCATATCAAGGACCATCGTATTTCTG 5084
399 erGluArgLeuSerGluGluLysThr...ProAspAsnLeuThrSer 414
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
5085 GAGAGAACTCTCGGACGACGAAAGCTGCGAGAAGCTGATAATCTCAAATCT 5134
```

MEDLINE REFERENCE AUTHORS TITLE JOURNAL	20330349 2 (bases 1 to 300650) Shirai, M. Direct Submission Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan. (E-mail: mshirai@go.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)	gene CDS	/db_xref="GI:8978644" /translation="MPDKKAQJTFSLPEVMSAIHQK1VALPDTDTVGVFLSVLYASEA EERLYALKDREPSKAFALYVNSIEDIENISGYPLSPAKKLAOLFPGAILTVVYKRRNP RFPKTLAFIVDHSVVREIVDHCGTLTGTSANLSEFPSSALTAEQIEAFDHLDCIF DGCSSHGLESTVANDPLIYRREGILSRSEVIENTAGTEAKIFHRTSHAFSKHIIKIVT KNEQIVSVFSLGSLDFKGVVCEHPKPNFTRLREALKKKTPSIVFIYDITSYDEL FPFLSPYIE" 3656..4387 /gene="CP10271" 3656..4387 /gene="CP10271" /codon_start=1 /transl_table=11 /product="lysophospholipase esterase" /protein_id="BAA98481.1" /db_xref="GI:8978645" /translation="MTDYSFRRKIGNIEAIECPGNQDDPIILCHGYSGLADNLTEF PSICSFKLRPTWIFPNGIIPLENDFGSRACFPILNVLLOELSRLLYANGVNGQEKY DELFDVLETPKAELEELIINLRPNYNEIIGFSGOAILATHLVLTQSNPYAGALIF AGARLFGNQWEEGLKQCAQVPFLQSHGYEDEILPYHLGAHLNLLLTQNGQVSVFHG GHEIPSVVFOKMVTVPNWDIPARG" complement(4359..5255) /gene="dnax_2" complement(4359..5255) /gene="dnax_2" /codon_start=1 /transl_table=11 /product="DNA polymerase III gamma and tau" /protein_id="BAA98482.1" /db_xref="GI:8978646" /translation="MHLEENQGWALLKRVYHQEVPAILLHGFTLPVLODKABOLA SEILLSSPGSEHKVSQKIPHDIVQFFPEKGRHSIDLPRGKIKQIVISPEAFANYKI YIIEADRMTLAIAFAFLKPEEPKHAIVILTAKVORLPKTIISRLSIFIERGEK ILCSEKSYFLRYAQCEIPIVTEVSQILIKESSTDQVLRDKVOREMVELLEYRDRY TLNGLKASALNYPFHVEKIELQLPLPLDKVLLIVESACRSNLNSSSASVLEWVAIQ LVSIQYKEKELVSVPQDLSN" complement(5246..5866) /gene="tdk" complement(5246..5866) /gene="tdk" /codon_start=1 /transl_table=11 /product="thymidylate kinase" /protein_id="BAA98483.1" /db_xref="GI:8978647" /translation="MFVIEGEGSGKSLAKALGDOLVAODRKVLLTRPQGGCLIGE RLRDLLEPHLELSCCELFELGSRAGHIOEVIIPALRGYIVICERFHDSTIVYQ CIAEGIGADFAVADLCQSKVGGPTFLPNFVLIDIPADIGIQKKHKQKPFKRPKPLS YHNRIRGFLSLASADPSRYLVLDARESASLIDKVMHLTQIGCT" complement(5871..8375) /gene="gyrA_1" complement(5871..8375) /gene="gyrA_1" /codon_start=1 /transl_table=11 /product="DNA gyrase subunit A" /protein_id="BAA98484.1" /db_xref="GI:8978648" /translation="MFNDEIIVPNLEEMKESVLRYSMSVSIISRALPDIDRGLKPS RVLQAMKQSLSPGAKHAKCAKTCGDTSGDYHPHGESVIPTILVRMAQNAMRYPL QVQGNFGISIDGPPAAHYTEARLTSHAMLMEDKDIIVDIPNYETKHEPVFP SKFNLALCNGSSGIAVGNATNIPPHNLGELFEATLLLANQASVDEILQVMPGDFP TGGIICGSEGIISTYTTGGRGKIKVRLARLHVEENDKHRESIIITEMPYNVKSRLIEQ IANLVNKTLAGISVRESKDGIRVYLEIKKGESSIIINRLYKFTDVOVTCANM LALNDLPTMSIHRMISAWIHRKVEIRRRIRYELNKAETRAHVEGLYKALSCLDA LVKTIRESNKEHAKERIIESFGTEPQALALELRULYOLTGLEAEKTOKEYLELUNK IAYYQVLSDEGLVKDIIIRNELQDLKHKHARRTTIEFDADDIIRDIEDIITNESVII TISGDDYVVRMPKVFKQORGGHGVTFDMKKGAGFKAVYSATTKDYLLFTNFGQ CIWLKWLPEGERAKGKPIINLEGIIPGEEALAINIKNFDNAGFLFLATKRGV KKVSLDASNPGRKGIIRLEDEGBELIAACHVIVSDEEKVMLFTHLGHAVRPHKVR PMGRTAGRVGSLNEEDKVVSCQIVTENSVLIVCDQGFGRSLVDFRETRNGGV GVRSILINERNQNLGAIPTVTDHDSILLMSQGOATIRNMQDVRVNRGSTRTOGVRVLVHL KEGDALVSMKLSNNENDDVLSGSEECSDTVSLR"
	On Sep 15, 2000 this sequence version replaced gi:6172290 gi:6172292 gi:6172294 gi:6172296 gi:6172326 gi:6172328 gi:6172330 gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172382 gi:6172384 gi:6172386 gi:6172388 gi:6172390 gi:6172392 gi:6172394 gi:6174666 gi:6174668 gi:6635174 gi:6635176 gi:6635178 gi:6635180 gi:8547433 gi:8547438 gi:8978640. AB033782-AB033785, AB033800-AB033815: Submitted (25-Oct-1999) AB038348-AB038349: Submitted (14-Feb-2000) AB036079-AB036082: Submitted (18-Dec-2000).		
FEATURES Source	Location/Qualifiers 1..300650 /organism="Chlamydomophila pneumoniae J138" /strain="J138" /db_xref="taxon:138677" /note="synonym:Chlamydia pneumoniae (strain J138)" 144..935 /gene="CP10267" 144..935 /gene="CP10267" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA98477.1" /db_xref="GI:8978641" /translation="MSIMSLSNKLNTALLNQEPVAVCLNWDPKYINQDKTKTACTVTL VIATMLITGVIVLLAMSGKSLVSTIIGTSVTLGTFALFGLVGLIKKSLAMI QYQYFOEVVTKGPEPISPKNDNVNHLTSCIPSPLODIESPSEASTPVSKLRACSG VAIVLGVTLGAVSVFFCTGYLOIALCVGACLTALFVCGLAGLRTLSLJAQGM YLYLYLVSSALRENETVQRNEINTYITTECRQOKREKALLE" 957..1343 /gene="CP10268" 957..1343 /gene="CP10268" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA98478.1" /db_xref="GI:8978642" /translation="MSQCQSSSTWEMKSFVFNWKNPTPLSPIPSEDEFTLAYEP FVLKTDENAGANPGTSTPNVNGIDDLNPLLPQNEQNANNPGTSGSNPTSLPA PERLPETENSOEQSQSONEDLIG" complement(1495..2472) /gene="CP10269" complement(1495..2472) /gene="CP10269" /codon_start=1 /transl_table=11 /product="dipeptidase" /protein_id="BAA98479.1" /db_xref="GI:8978643" /translation="MTIDMHCDDLHSHPHFCRDKPAVRCSPEQLLSGGVQRQVAFPP HSRPECNDKNSLFFSPNOYPDIGLLSYEEENGSSOKKSLISIRJENASALGD DTAPLCTLLAKLHLTKGPGPLAYLGIYVWKGUNRPGCGTEAPKRI SDCVKLLDIWEL GVYDLSHCSKDLAEDILDTADKLPNLAV IASHNSFRSLDHRNKLVDNAKEIVRR KGVTLNLRVYSGVSLGDLKHLVAENLSLIVLSGDFFFYANEDENEFNFCSS AEAPVNLQIHRIFSKAEPSSILSSRAEKLKQVIVEQVNPKITDKL" complement(2487..3347) /gene="ywlC" complement(2487..3347) /gene="ywlC" /codon_start=1 /transl_table=11 /product="SuA5 superfamily-related protein" /protein_id="BAA98480.1"	gene CDS	/db_xref="GI:8978644" /translation="MPDKKAQJTFSLPEVMSAIHQK1VALPDTDTVGVFLSVLYASEA EERLYALKDREPSKAFALYVNSIEDIENISGYPLSPAKKLAOLFPGAILTVVYKRRNP RFPKTLAFIVDHSVVREIVDHCGTLTGTSANLSEFPSSALTAEQIEAFDHLDCIF DGCSSHGLESTVANDPLIYRREGILSRSEVIENTAGTEAKIFHRTSHAFSKHIIKIVT KNEQIVSVFSLGSLDFKGVVCEHPKPNFTRLREALKKKTPSIVFIYDITSYDEL FPFLSPYIE" 3656..4387 /gene="CP10271" 3656..4387 /gene="CP10271" /codon_start=1 /transl_table=11 /product="lysophospholipase esterase" /protein_id="BAA98481.1" /db_xref="GI:8978645" /translation="MTDYSFRRKIGNIEAIECPGNQDDPIILCHGYSGLADNLTEF PSICSFKLRPTWIFPNGIIPLENDFGSRACFPILNVLLOELSRLLYANGVNGQEKY DELFDVLETPKAELEELIINLRPNYNEIIGFSGOAILATHLVLTQSNPYAGALIF AGARLFGNQWEEGLKQCAQVPFLQSHGYEDEILPYHLGAHLNLLLTQNGQVSVFHG GHEIPSVVFOKMVTVPNWDIPARG" complement(4359..5255) /gene="dnax_2" complement(4359..5255) /gene="dnax_2" /codon_start=1 /transl_table=11 /product="DNA polymerase III gamma and tau" /protein_id="BAA98482.1" /db_xref="GI:8978646" /translation="MHLEENQGWALLKRVYHQEVPAILLHGFTLPVLODKABOLA SEILLSSPGSEHKVSQKIPHDIVQFFPEKGRHSIDLPRGKIKQIVISPEAFANYKI YIIEADRMTLAIAFAFLKPEEPKHAIVILTAKVORLPKTIISRLSIFIERGEK ILCSEKSYFLRYAQCEIPIVTEVSQILIKESSTDQVLRDKVOREMVELLEYRDRY TLNGLKASALNYPFHVEKIELQLPLPLDKVLLIVESACRSNLNSSSASVLEWVAIQ LVSIQYKEKELVSVPQDLSN" complement(5246..5866) /gene="tdk" complement(5246..5866) /gene="tdk" /codon_start=1 /transl_table=11 /product="thymidylate kinase" /protein_id="BAA98483.1" /db_xref="GI:8978647" /translation="MFVIEGEGSGKSLAKALGDOLVAODRKVLLTRPQGGCLIGE RLRDLLEPHLELSCCELFELGSRAGHIOEVIIPALRGYIVICERFHDSTIVYQ CIAEGIGADFAVADLCQSKVGGPTFLPNFVLIDIPADIGIQKKHKQKPFKRPKPLS YHNRIRGFLSLASADPSRYLVLDARESASLIDKVMHLTQIGCT" complement(5871..8375) /gene="gyrA_1" complement(5871..8375) /gene="gyrA_1" /codon_start=1 /transl_table=11 /product="DNA gyrase subunit A" /protein_id="BAA98484.1" /db_xref="GI:8978648" /translation="MFNDEIIVPNLEEMKESVLRYSMSVSIISRALPDIDRGLKPS RVLQAMKQSLSPGAKHAKCAKTCGDTSGDYHPHGESVIPTILVRMAQNAMRYPL QVQGNFGISIDGPPAAHYTEARLTSHAMLMEDKDIIVDIPNYETKHEPVFP SKFNLALCNGSSGIAVGNATNIPPHNLGELFEATLLLANQASVDEILQVMPGDFP TGGIICGSEGIISTYTTGGRGKIKVRLARLHVEENDKHRESIIITEMPYNVKSRLIEQ IANLVNKTLAGISVRESKDGIRVYLEIKKGESSIIINRLYKFTDVOVTCANM LALNDLPTMSIHRMISAWIHRKVEIRRRIRYELNKAETRAHVEGLYKALSCLDA LVKTIRESNKEHAKERIIESFGTEPQALALELRULYOLTGLEAEKTOKEYLELUNK IAYYQVLSDEGLVKDIIIRNELQDLKHKHARRTTIEFDADDIIRDIEDIITNESVII TISGDDYVVRMPKVFKQORGGHGVTFDMKKGAGFKAVYSATTKDYLLFTNFGQ CIWLKWLPEGERAKGKPIINLEGIIPGEEALAINIKNFDNAGFLFLATKRGV KKVSLDASNPGRKGIIRLEDEGBELIAACHVIVSDEEKVMLFTHLGHAVRPHKVR PMGRTAGRVGSLNEEDKVVSCQIVTENSVLIVCDQGFGRSLVDFRETRNGGV GVRSILINERNQNLGAIPTVTDHDSILLMSQGOATIRNMQDVRVNRGSTRTOGVRVLVHL KEGDALVSMKLSNNENDDVLSGSEECSDTVSLR"

1773 AAGGGCAGCCCGCCAGCAATCTATCTATTCTATGATCCGATTGCATCTA 1724
358 luSerSerSerLysGluSerProLeuProSerSerSerLeuGlnAlaSerVal 374
1723 ACACCAAGAGGCTTCAGACGTTCTG..... 1698
375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
1697 ACCATCAACCAACCGGATAGCACTCGCTTTA.....GATTA 1660
391 aAsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGluL 407
1659 TTCAGGAACGATTGTTATTTCTGGGAAGANGCTCTCTGAGATGAAGCGA 1610
407 ystThrProAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
1609 AAGCTGCTGATACTTCACATCTATATTAAGCAACCATTTGGCTCTAGCC 1560
424 SerGlyArgLeuValLeuLysAspAlaValAlaValLeuSerAlaProSerLe 440
1559 TCTGGAACCTTACCACTCAAGGAAGATGTCGAGCTTAGATGTCATGTTT 1510
440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuL 457
1509 CACACAGACTGAAGGCTCTACATCTCTCATGCAACCAAGCAACAGCTCA 1460
457 ystThrSerSerAsp...LeuLysIleAlaThrLeuSerIleProLeuHis 472
1459 AAGCAGATACCTAAGCTATCAGCTCTTACCAAACTTGTCTGTTGATCTTCT 1410
473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485
1409 GCCTTAGAGGAATAGAGTGTCTCCATTGAACACAGCAGGAGCCCAACAA 1360
486AsnLeuSerIleGlnIysIlePheLeuSerAsnSerGlyAspG 500
1359 AACTATAACTCTAACCTCTCTCTGTTTTCAGAGATAGTAGGCGC.... 1314
500 luAsnPheThrGluAsnValGluLeuLeuSerLysGluGlnAsnAlaIle 516
1313 ..AATTTTATGAAGCCATACGATACCAACCAAGCCCTTCAGCGACGCTTG 1266
517 ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533
1265 GTGGATTTCACCTGCTGCTACTGCTGCTAGGATATTTATATC...GATGC 1219
533 yAsnLeuSerSer.....HisPheGlyTyrGlnG 543
1218 GCCTTCACCTCTCCAGTACAACTCCAGAACCTCTACGGGTATCAGG 1169
543 lyAspTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeuIle 559
1168 GACATTGGCAAGCCACTTGGCGACACATCACTACCAAAATCAGGAAC 1119
560 AlaAsnTrpThrProLysAsnTyrValProHisProGluArgGlnSerTh 576
1118 ATGACTTGGGTAAGTACGGGCTACACCCCTAACTCTGAGCGGTAGAGGCTC 1069
576 rLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG 593
1068 CGTAGTTCCCGATTCTATATGGGCATCTTTACTGACATTCCGACTCTAC 1019
593 lnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThr 609
1018 AGCAGATCATGACATCTCAAGCGCAATAGTATCTATCAGCAACGAGGACTC 969
610 TrpClySerAlaValSerAsnLeuPheThrAlaHisAspSerSerGlyLy 626
968 TGGGCATCAGGAAGTGGCAATTTCTTCAT...AAGGATAAATCAGGAAC 922
626 sProlleAspAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleS 643
921 TAAC...CAGCATTCGACATCAAAAGCTACGGCTATATTGTTGGAGAA 875

643 erThrHisSerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeu 659
874 GTCTCAAGATTTTCTGAAATATATCTTACGTAGCTTTCTGCCAGCTC 825
660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTy 676
824 TTCGGTAAAGATAAAGACCTGTTTATAGTTGAAATACCTCTCATACCTA 775
676 rIleAlaThrValGlnAlaGln..... 683
774 TTTAGCGTCGTATACCTGCAACATCAGCAGATTCTCTAGGAGGACTCCCA 725
684LeuAlaThrSerLeuMetLys..... 690
724 TGCCTCATTTTGAAGTATCACCGACATGCTGAAAGATATCTCTCATTT 675
691 IleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeuLysThrLy 707
674 TTCAATGCCCAGCTAAGCTACAGCTACACTAAATAATGATATGGATACTCG 625
707 sTyrArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAlav 724
624 CTACTCTTCCTAT...CCTGAAGCTCAAGGCTCTTGGACCAATAAATCTG 578
724 alSerGlyGluValCysAlaSerIleProIle...ValSerAsnGlySer 739
577 GGCTCTAGACCTCGGAGGATCTCTGGCTCTATATCTCCCTAAAGAAGCA 528
740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerG 756
527 CCGTTCTCCAGGGATATTTCCCTCTTAAAGTTCCAGGCAGCTACAG 478
756 yThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerPheSerA 773
477 CCGCAACAAACTTTAAAGAGAGTGGCGCTGAAGCCGCTGTTTGATG 428
773 laSerSerPheArgAsnIleSerLeuProIleGlyIleThrPheGluLys 789
427 ATGAGACCTAGTCACTGCTCTATCCCTGTCGCCATTTCGGTTAGAAAA 378
790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyrIleG 806
377 ATCTCCGAAGATGAAAAATAATTTCCGAGATTCTCTAGCCTACATTGG 328
806 nAspLeuLysArgAspValGluSerGlyProValValLeuLeuLysAsnA 823
327 TGATGTGTATCGTAAAAATCCCGCTTCGCGTACTCTCTAATGTCAGTG 278
823 laValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAlaTyrMet 839
277 GAGCCTCTGGACTTCGCTATGTAAAAACCTCGCACGACGAGCCCTCTTA 228
840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeuLe 855
227 GCAAGTCTGGAAGCCACTGACTCTCTCCCTCATGTAGAACTCTCTGG 178
855 uAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerLeuAspL 872
177 GGAAGCTGCTTATGAGCTTCGTGCTCAGCACACATCTACAATGTAGATT 128
872 euGlyThrThrTyrArgPhe 878
127 GTGGGCTAAGATACCTCATTC 108

seq_name: gb_bai:AE001627

seq_documentation_block:
LOCUS AE001627 15068 bp DNA BCT 01-DEC-2000
DEFINITION Chlamydia pneumoniae section 43 of 103 of the complete genome.
ACCESSION AE001627 AE001363
VERSION AE001627.1 GI:4376721
KEYWORDS

FAADKNSLSCTIALIDEGSFYENHNLKSASTYPLLELTAGANGTITLGAISTLTL
QEPETHYCGNOLWSANATSKIGSINWTRTGYIPSPERKSNLPLNSLWGNFIDIR
SINLOIETSGEPEFRELWSLSTANFETGRDMSFTRGFRHISGGYALGATITPAED
QITAFOLFARDNRHITGNHGDYTGASLYFHHTGLEDFDIANFLWAGKATRAPWLS
ISQIPLFSYLFADNYHTDMSITITKISGNSTISITFSNAKLGGAIIYSSAAISGNTGL
LKEVPVKQYIYVAHOODEFVHAEGRAFNFKNSELINVEIPIGVTFERDSKSEKGYD
LTMWILIDAYRRNPCKOTSLIASDMMVMAYGTNLAROGFSYVAANHFQVNPMEIFQ
FAEYRSSRNYNTNLSKFCF"

gene

12172..14964

/gene="pmp_8"

/note="Cp0046"

12172..14964

/gene="pmp_8"

/codon_start=1

/translation=1

/product="Polymorphic Outer Membrane Protein G Family"

/protein_id="AADI8590.1"

/db_xref="GI:4376729"

/translation="MKPLHLKLLISSTLVTPILLISVIATYGADASI.SPTDSFDGAGST
FPTKSTADANGTNYVLNGYVINDAGKGTALTGCTFTTGDLTFTGCKGYSFSENTVD
AGSNAGAASSTADKALTFTGFSNLSFTAPAGTTVASGKSTLSSAGALNLDNGTILF
QVNSFNANNGNGAITTKLISIGNSTISITFSNAKLGGAIIYSSAAISGNTGL
VFMNKGTEGGALGFEASSITONSLIFFSGNTATDAAGKGAIIYCEKGETPTLI
SGNKLTAFAENSSVTOGCAICAHGLDLSAAGPTLFSNNRCGNTAAGKGAIIADSGS
LTSANOGDITFLONTLTSTAPTSTRNAIYLGSSAKITNLRAAGOSIYFYDPIASN
TSGASDLVLTINQPOSNSPLDYSGTIVFSGEKLSDAEKADANFTSILKQPLALASGTL
ALKGNELDVNGFTQTEGSLMQPGTKLKADTEAISLTKLVLDLSALEGNKSVSIE
AGAKNTITLSPVQSSNGFYESHTINOATQPLVFTTAASDIYIDALLTSPV
TQPPHYQGHWAETADWTAKSGTWTWTTGPNPERRASVVPDSLWASSETDIR
TLQOIMSNASIIYOORCLWASGTANPHKDKSGTNOAFRHKSIVYIVGSAEDFSEN
IFSVAFOLFCKDKDLFIVENTSHYLSLYLQHRFLGGLPMPSEFSITDMLKDPL
ILNQLSISYTKNDMDTITSTYPAQGSNTNSGALGGLSGLALYLPKEAPFQGYFP
FLKQVAYSROQNEKESGAARFDDGLVNCISIPVGRLEKISEDEKNFEISLAYI
GDVVRKNPRTSLMVSASWTSLCKNLARQAFLASAGSHLTLSPHVLSGEAAVELR
GSAHYNVDCGLRYSF"

BASE COUNT 4358 a 3550 c 2979 g 4181 t

ORIGIN

alignment_scores:

Quality: 915.50

Length: 840

Ratio: 1.705

Gaps: 26

Percent Similarity: 63.929 Percent Identity: 30.833

alignment_block:

US-09-677-752-4 x AE001627 ..

Align seg 1/1 to: AE001627 from: 1 to: 15068

98 LeuSerIleGlnAsnPhaArgPheLeuSerPhe.....Th 109

12526 CTAACATTCACAGATTTCCTTACCTTCCTTCATTGCGCTCCTGGAAC 12575

109 rAspCysSerSerlyGluSerProSerPheIleHisGlnIlyAsnG 126

12576 TACAGTTGCTTCAGGAAAGTACTTAAAGT.....TCTGCAG 12613

126 lyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg..... 140

12614 GAGCCTTAATCTTACCGATAATGGAACGATCTCTTTACCCAAACGTC 12663

141 ...AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSe 156

12664 TCCAAATGAAGCTAATAACAATGGCGGAGCGATCACCACAAAACCTCTTC 12713

156 rLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerlySG 173

12714 TATTTCTGGGAATACCTCTTCTATATAACCTTCACCTAGTAATAGCGCAAAA 12763

173 lyAsnGlyGlyAlaIle...GlnAlaGlnThrPheSerLeuSerArgAsn 188

12764 AATTAGGTGGAGCCATCTATAGCTCTCGGCTGCAAGTATTTCAGGAAC 12813

189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAl 205
12814 ACCGCCAGCTTAGCTTTTATGAATATAAGGAGAACTGGGGTGGGC 12863
205 atleCysCysSerAsnLeuIleCysSerGlyAsnVal.....AsnP 219
12864 TCTG.....GGCTTTGAAGCCAGCTCCTCGATTACTCAAAATAGCT 12904
219 rLeuPhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
12905 CCCTTTCTCTCTGGAACACACTGCAACAGATGCTCGAGCAAGGGCGG 12954
232 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSe 248
12955 GCCATTATTGT.....GAAAAACAGGAGAGACTCTTACTTACT 12995
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluL 265
12996 TATCTCTGAAAATAAAGTCTGACCTTCGCCGAGAACTCTTCAGTAAC 13045
265 ysGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyPro 281
13046 AAGCGGAGCAATCTCTGCCCATGGTCTAGATCTTTCCGCTGCTGCCCT 13095
282 ValSerPheIleAsnAsn.....SerAlaLysIleGlyGly 293
13096 ACCTTATTTCAANTATATAGATCGGGAACACAGCTGCGGCAAGGGCGG 13145
293 yAlaIleAlaIleGlnSerGlySerLeuSerIleLeuAlaGlyGluG 310
13146 CGCTATTGCAATTGCCGACTCTGGATCTTTAAGTCTCTCTGCAAACTCA 13195
310 lySerValLeuPheGlnAsnAsnSer...GlnArgThrSerAspGlnGly 325
13196 GAGACATCAGCTTCTTGGCAACACTTAACCTCAACCTCCGCCCAACA 13245
326 LeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLe 342
13246 TCGACACGGAATGCTATCTACCTGGGATCGTCAGCAAAATTAGCACTT 13295
342 uGluAlaArgAsnGly...AspIleLeuPhePheAspProIleValGlnG 358
13296 AAGGCGACCCCAAGGCCAACTCTATCTATTCTATGATCCGATTCGATCA 13345
358 luSerSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerVal 374
13346 ACACACAGAGCTTTCAGAGCTTCTG..... 13371
375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
13372 ACCATCAACCAACCGGATAGCACTCGCCTTTA.....GATTA 13409
391 aAsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGluL 407
13410 TTCAGGAACGATCTTATTTCTGGGAAAACCTCTCTGCAGATGAAGCGA 13459
407 ysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGluLeuLys 423
13460 AAGCTGCTGATAACTTCACATCTATATTAAGCAACCATTTGGCTCTAC 13509
424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLe 440
13510 TCTGGAACCTTAGCACTCAAGGAAATGTCGAGTTAGATGATCAATGGTTT 13559
440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuL 457
13560 CACACACTGAAGCTCTACACTCTCTCATCATCAACAGGACAAAGCTCA 13609
457 ysThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
13610 AAGCAGATACCTGAAGCTATCAGTCTTACCAAACTTGTGTTGATCTTCT 13659
473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485

```

1239 TTCCAGGAACGATTGTATTTCTGGGAAAAAGCTCTCTGCAGATGAAGCGA 1288
407 yThrProAspAsnLeuThrSerGluGlnGlnProIleGluLeuLys 423
1289 AGCTGCTGATACCTTACATCTATATTAAAGCAACCATTTGGCTCTAGCC 1338
424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLe 440
1339 TCTGGAACCTTAGCACTCAAGGAATGTCGAGTTAGATGTCATGTTT 1388
440 uSerGlnAspProGlnAlaLeuLeuLeuMetGluAlaGlyThrSerLeuL 457
1389 CACACAGCTGAAGGCTCTACACTCTCTCATGCAACCAAGCAACAAAGCTCA 1438
457 yThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
1439 AGCAGATACCTGAGCTATCAGTCTTACCAAACTTGTCTGTGATCTTCT 1488
473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485
1489 GCCTTAGAGGGAATTAAGAGTGTGTCATTGAACACGAGGAGCAACAA 1538
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspG 500
1539 AACTATAACTCTAACCTCTCTCTCTGTTTTCGAAGATAGTAGCGGC.... 1584
500 LuAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsnAsnIle 516
1585 ..AATTTTATGAAGCATACGATAAACCAAGCCTTCACGCGCCTTTG 1632
517 proLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533
1633 GTGGTATTCTGCTGCTACTGCTGCTAGCGATATTTATATC...GATGC 1679
533 yAsnLeuSerSer.....HisPheGlyTyrGlnG 543
1680 GCTTCTCACTTCTCCAGTACAACTCCAGAACCTCATTTACGGGTATCAGG 1729
543 LyAspTyrThrPheSerTrpLysAspSerAspGluGlyHisSerLeuile 559
1730 GACATTGGGAAGCCACTTGGCGACACATCACTGCAAAATCAGGAAC 1779
560 AlaAsnTrpThrProLysAsnTyrValProHisProGluArgGlnSerTh 576
1780 ATGACTTGGGTAACTTCGGGCTACACCTTAATCTGAGCGTAGAGCTTC 1829
576 rLeuValAlaAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG 593
1830 CGTAGTTCCCGATTCAATATATGGCATCTCTTACTGACATTCGCACCTTAC 1879
593 InSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThr 609
1880 AGCAGATCATGACATCTCAAGCGAATAGTATCTATCAGCAACGAGGAGCTC 1929
610 TrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLys 626
1930 TGGCATCAGGAACCTGCGAATTTCTTCCAT...AAGGATAAATCAGGAAC 1976
626 sProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleS 643
1977 TAAC...CAAGCATTCGACATTAAGGCTACGGCTATATTTGTTGGAGAA 2023
643 erThrHisSerLeuAspHisSerPheCysLeuAlaAlaGlyGlnLeu 659
2024 GTGCTGAAGATTTTCTGAAATATCTTCACTGAGTGTAGCTTTCTGCCAGCTC 2073
660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSerTy 676
2074 TTCGTAAGATAAGACCTGTTTATAGTTGAAATACCTCTCTCATAACTA 2123
676 rIleAlaThrValGlnAlaGln..... 683
2124 TTACGCTCGCTATACCTGCAACATCGAGCATTCCTAGGAGGACTTCCCA 2173

```

114 sGluSerSerProSerIleIleHis.....GlnLysA 125
 125 snGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArgAsn 141
 2490TCTCTAGTGAATAGTGAAGAGGGCTATGAATCCG 2526
 125 snGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArgAsn 141
 2527 GAGGAGCATTAAACTTACCAATTAATCCAGTATTCGTGATCAGAAC 2576
 142 HisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLeuG1 158
 2577 TATTCCTCAGATGTGGAGCACTCTCTGCAAGCTTTTCTCTAAC 2626
 158 nHisAsnTyLeuPheThrAlaPheGluAsnSerSerLysGlyAsnG 175
 2627 CGGCTCCAGCAAGAAATCAGCTTACCACCTAATCTTACTCCGAAAAAG 2676
 175 lyGlyAlaIleGlnAlaGlnThrPheSer...LeuSerArgAsnValSer 190
 2677 GTGAGGCGATTGCTGTACGGGAATAGCTCATCTTCGGACAACCAAGGC 2726
 191 ProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCy 207
 2727 ACAATCAGATTTCTGGGAACACTGCTGTGAATTCGGGGGACAGTATA 2776
 207 scys...SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePhe 223
 2777 TTCAGAAGCTTCTATGACGATTCAGGTAAAC...AACCACGTTGCTTTTA 2823
 223 hrGlyAsnSerAlaThrAsn.....GlyGlyAlaIleCys 234
 2824 GCACAAATGCTGTTCGGGTCTATCTGATGTTGCGGTGGAGCTATCCAT 2873
 235 CysIleSerAspLeuAsnThrSerGluLysGlySerLeuAlaCy 251
 2874 TGT.....ACCAAAACAGGTTACGACCCACCTTACTATAAGAGA 2914
 251 sAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlyGlyAla 268
 2915 TAACAAAGCTTGTATTTTTCAGGAAAACTTCTTCAGCAAAAGGTGAG 2964
 268 laIleTyAlaLysHisMetValLeuArgTyraAsnGlyProValSerPhe 284
 2965 CGATTTACCCGATAAATCATATTGACTTCTGGTGGGCTACGGCATTT 3014
 285 IleAsnAsn.....SerAlaLysIleGlyAlaIleAlaI 297
 3015 ATCAATAACAAAGTTACCCATGCTACACCTAAGGGTGGAGCTATTGGTAT 3064
 297 eGlnSerGlyGlySerLeuSerIleLeuAlaGlyGlySerValLeuP 314
 3065 TGCTGCCAATGGAGATGTAGTTAAACCGCTGAACATGGGGATATTACTT 3114
 314 heGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeuVal...ArgAsn 329
 3115 TTGATAATAACCTGATGGCCACACAGACAACTGCTACAATAAAGAAAT 3164
 330 AlaIleTyLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArgAs 346
 3165 GCATTAACATTTAAGGCAATGGTAATTCGTCACTTACGTACGGGTC 3214
 346 nGly...AspIleLeuPhePheAspProIleValGlnGlySerSerL 362
 3215 TGGAAAGACCATTTCTTCTATGATCCTCATCAGATTGAAGGTAATGCTG 3264
 362 ysGluSerProLeuProSerSerLeuGlnAlaSerValThrSerProThr 378
 3265 CTGATCTCTCACTTTGAATTAAGCTGAGGGTGATAAAGC..... 3305
 379 ProAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArgSera 395
 3306TATTAATGGAAGAAAT 3319

395 IlePheSerSerGluArgLeuSerGluGluGlu...LysThrProAspA 411
 3320 TATTTTTTTCAGGAGAAAAGCTCACTAAGAACACAGCTGCTTTCGGGATA 3369
 411 snLeuThrSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeu 427
 3370 ACCTAAAGACAAATTTACAGCCTTATCACTTTAGCTGCTGGTGAACCT 3419
 428 ValLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspPr 444
 3420 GTGTTACGCGAGCGGTGGAAGTGAAGCAAAACACAGCTGCTGCAACACG 3469
 444 oGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuLysThr...SerS 460
 3470 AGGATCTTTGATTCTGTGATGATGATGATGATGATGATGATGATGATG 3519
 460 erAspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThr 476
 3520 AAGATGCTACACTGACCAATCTGGCTATTATCCGAATACCTTAGATGG 3569
 477 GluLysSerValThrIleHisAla.....ProAsnLeuSerIle 489
 3570 AAAAAATTCGCGTAGTGCATGCGTTCGCTGGGGAAGATGTGACTTT 3619
 489 eGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTyGluAsn. 505
 3620 ATCAGTGCTTATTCGCGCTTATTCATCTACAGGCAAGCTTTATGAAAC 3669
 506ValGluLeuLeuSerLys 511
 3670 ATAAGCTAAATGATACGTTAGCTTTAGGAGGAATTCACACTTTCTGGGAA 3719
 512 GluGlnAsnAsnIleProLeuLeuThrLeuSerLysGluGlnSerHisLe 528
 3720 GGTTCG.....GTGACAAACCAACCAACGCTGCTACTCATGT 3754
 528 uHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrglnGlyAsp 545
 3755 T.....GTTGGTGTCTGAAACCCACTATGTTTATCAAGGAACT 3795
 545 rPheThrPheSerTrp...LysAspSerAsp.....GluGlyHisSer 557
 3796 GGCTGTGCTGGGTCAAGATATAAATCTGATCTCAACACCAACAA 3845
 558 LeuIleAlaAsnTrpThrProLysAsnTyValProHisProGluArgG1 574
 3846 GCATCTTTTACCTGGAAATAAACAGGATATGTTCCAAATCCTGAACGTCG 3895
 574 nSerThrLeuValAlaAsnThrLeuTrpAsnThrTySerAspMetGlnA 591
 3896 TGCTCCGCTAGTACTCAATAGCTTTGGGATCCTTTATAGATTACGTT 3945
 591 laValGlnSerMet.....IleAsnThrIleAlaHisGlyGly 603
 3946 CTATTCAAGATGCTTGGAAACGATGTTGATAGTATTCTTGAGACACGT 3995
 604 AlaTyLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheTyAl 620
 3996 CGT.....GTTTGTGGCTCTCTGGAAATGGGAACTTCTTC..... 4031
 620 aHisAspSerSerGlyLysProIleAspAsnTrpHisHisArgSerLeuG 637
 4032 .CATAAGATCGGAATGCTCAAAATCCAAATCCCGTCATATCAGTTCGG 4080
 637 lyTyLeuPheGlyIleSerThrHisSerLeuAspHisSerPheCys 653
 4081 GATATGTGTTAGGACCAACCAATACCTCGAGAGAGGATTCTCTTAGT 4130
 654 LeuAlaAlaGlyGlnLeuLysSerSerSerSerPheIleThrSe 670
 4131 GTGCTTTCTGCTCAGTTATTTGCAAAAGATAAGACCTACCTTGTAAACAA 4180
 670 rThrGluThrThrSerTyIleAlaThrValGlnAlaGlnLeuAlaThrS 687

1577 GG.....AATGCTTATGAAATCAGCAC...TTAGGAAAAAAGCTCAAGAC 1611

515AsnIleProLeuLeuThrLeuSerLysGluInSerHis 527

1618 TTTCATTGTGCAGCTCTCTGCTCTGGGTACTCGCAACAACACACAGATGT 1667

527 sleuHisLeuProAspGlyAsnLeuSerSerHisHisPheGlyTyrGlnGlyA 544

1668 TCCAGCGGTTCTCTACAGTAGCAACCTCTACGCACATATGGGTATCAAGGTA 1717

544 sPrThrPheSerTrpLysasp.....SerAspGluGlyHis 556

1718 CTTGGGGAAATGACTTGGGTGATGATACCGCAAGCACTCCAAAGACTAAG 1767

557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluAr 573

1768 ACAGCGACATTAGCTTGCACCAATACAGGCTACCTCCGAATCCTGAGCG 1817

573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590

1818 TCAAGGACCTTTAGTTCTCTAATAGCCTTTGGGGATCTTTTCAGACATCC 1867

590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606

1868 AAGCATTCAAGCTGCTCATAGAGAAAGTGCTTTGACTCTTTGTTCAGAT 1917

607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623

1918 CNGAGCTCTCGGCTCGGGAGTCGCCAAT...TTCITAGATAAAGATAA 1964

623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640

1965 GAAGAGGGAAAAA...CGCAATAACCGTCATAAATCTGGTGGATATGCTA 2011

640 heGlyIleSerThrHisSerLeuAspAspHisHisSerPheCysLeuAlaAla 656

2012 TCGAGGTCGACGGCAACTTGTCTGAAACTTAATACGTTTGGCCTTT 2061

657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673

2062 TGCCRACTCTTTGGTAGCGATAAAGATTCTTCTAGTCGCTAAAATCATAC 2111

673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688

2112 TGATACCTATGACGAGGCGCTTCTATATCCACACATTACAGAATGTAGTG 2161

688 688

2162 GGTTCATAGGTTGTCTCTTAGATAAACTTCCTGGCTCTTGGAGTCATAAA 2211

689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704

2212 CCCCTCGTTTAAAGAGGCGAGCTCGCTTATAGCCACGTCAGTATGATCT 2261

704 ubYsThrLysTyrArgSerPheSerLysGluGlyPheGlySerThrPhis 721

2262 GAAGACAAAGTACTGCGGTAT...CCTGAGGTGAAGGTTCTTGTGGGGA 2308

721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737

2309 ATAATCGTTTAAACATGATGTGGGAGCTCTCTTCATCTTATCCCTCAA 2358

738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754

2359 TACCTCGATTGTTTGTATACCTATGCTCCATACATCAAACTGAATCTGAC 2408

754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyGluIleArgSerP 771

2409 CTATATACGTCAGGACAGCTTCTCGGAGAAAGGTACAGAAGGAAGATCTT 2458

771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787

2459 TTCTACAGACACACCTCTTCAATTTATCTTTCCTTATAGGGGTCAAGTT 2508

7634	TCACGGAGGCTTGTGGAAACTTTATGCCAATAACATTACGACGACGCCC	7683			
838	TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeu...GlnTh	853	:::		:::
7684	TTGCAAGTGCGTGACGACACTCAGCGCTTCTCTCCATCTTGAAGCT	7733	:::		:::
853	rLeuLeuAsnValSerCysValleuArgGlyGlnSerHisSerTyrSerL	870			
7734	CCTCGGCACATGTTCTTTGAAGTTCTGTCGATCCTCACGGATTTAATG	7783			
870	euaSpIeuGlyThrThrTyArgphe	878			
7784	TAGATCTTGGGGTAGTTCCAATTC	7809			
seq_name: gb_ba2:CPN133034					
seq_documentation_block:					
LOCUS	CPN133034	17280 bp	DNA	BCT	14-NOV-2000
DEFINITION	Chlamydia pneumoniae GA1-protein family cluster A, genes omp12 (partial), omp11, omp10, omp5, omp4, omp13 and omp14 (partial) and ORF1, strain VR1310.				
ACCESSION	AJ133034				
VERSION	AJ133034.1	GI:4455881			
KEYWORDS	hypothetical protein; omp10 gene; omp11 gene; omp12 gene; omp13 gene; omp14 gene; omp4 gene; omp5 gene; ORF1; outer membrane protein; outer membrane protein 11; outer membrane protein 12; outer membrane protein 14; outer membrane protein 4; outer membrane protein 5.				
SOURCE	Chlamydophila pneumoniae.				
ORGANISM	Chlamydophila pneumoniae				
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.				
AUTHORS	Hjerno,K., Boesen,T., Daugaard,L., Knudsen,K., Madsen,A., Christiansen,G. and Birkelund,S.				
TITLE	Chlamydia proteins containing the GCAL-repeat belong to a subfamily of autotransporting pathogenicity factors				
JOURNAL	Unpublished				
REFERENCE	1 (bases 1 to 17280)				
AUTHORS	Boesen,T.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-JAN-1999) Boesen T., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, DK-8000 Aarhus, DENMARK				
FEATURES	Location/Qualifiers				
source	1..17280				
organism	"Chlamydophila pneumoniae"				
strain	"VR1310"				
db_xref	taxon:83558"				
gene	1..841				
cds	<1..841				
note	"omp12"				
transl_table	2				
product	"outer membrane protein 12"				
protein_id	"CAB37067.1"				
db_xref	"GI:4455882"				
translation	"EDNIYRHNSGGVALGITATPAEDQLTFAPCLFARDRNHTIG NKGDTYGASLYFHTEGLEDFIANFLMGKATRAPWLISEISQIPLSFDAKFSLYHTD NMKTYTDNSITIKGSWRNFADGLGALPLPVISVPYLKEVEFPVKVOYIYAHOHD FYERHAEGRAKNSELINVELPIGVTFERDSKSEKGYDLTLMLYILDAYERNPKCOTS LIASDNWAYGTNLARQGFSVRANHFVNPHMIFPGQAFEFVRSRRNYNTNLASK FCF"				
sig_peptide	885..967				
gene	885..3677				
cds	885..3677				
note	"putative"				


```
ILCKSTFTSYLRYAQCELPVTEVSQIIKSSSETDXQVLRDKVOREMEVLELLYRDYV
TLNIGLKASALNYPHEVKEIILQPLLPIDKVLVLLIVESACRLNNSSSAAASVLEWVAIQ
LVSLYKREKELVSYSPQDLN"
/sequence="tdk"
complement(5246..5866)
/sequence="tdk"
complement(5246..5866)
/codon_start=1
/transl_table=11
/product="thymidylate kinase"
/protein_id="BAA98483.1"
/db_xref="GI:8978647"
/translation="MFIVIEGEGSGKSSLAALGDOLVAODRKVLLITREPCCGLIGE
RLRLDILPEPHLELSRCELEFLGSRQAHIQOEVIIPALRDGYIVICERFHDSTIVYO
GIAEGLGADFVADLCSSVGPFPFNPVLLDIPADIGQIRKQKQVDFKFKKPLS
YHNRIRGFLSLASADSPSLVLDARESLASLDKVLMLTQLGLT"
complement(5871..8375)
/sequence="gyrA_1"
complement(5871..8375)
/sequence="gyrA_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit A"
/protein_id="BAA98484.1"
/db_xref="GI:8978648"
/translation="MFNKEDEIIVPKNLEEMKESYLRYSMVSIISRALPDIDRGLKPS
QRRVLYAMKQLSLSPGAKRHKCAKIGDTSYDHPHGESVIPTLVRMAQNMARYPL
VGGGNGFSIDGDPAAAMRYTEARLTHSAMVLMEDLKDIDVIDVNPYDETHIEPVFP
SKFPLNLCGSSGIVAGWATNIPHNLCGLIEATLLLANPQASVDEILQVMPGDDFP
TGICGSEGIRSYTHITGRGKIKYPARLHVEENEDKHRESIIITPEYVNVKSRLLIEQ
IANVYNKTLGASIDVSDSDKIRVVLLEIKKGESSEIIINRLKFTDQVTFGANM
LALDKLPRTSHIRSIAMIRHRKEVIRRRTRYELNKAETRAHVELEGYALKSLCDA
LVKTRRESGHEKARIETSEFTEPOALAILLELYQLTGLAEAKTOKEYEELLNK
TAYQVYKRWPKVYKFEKQRRGGHGVTFGDMKKGAGFLKAVYSFTKYDLIFTNFGO
TISGDVYKRWPKVYKFEKQRRGGHGVTFGDMKKGAGFLKAVYSFTKYDLIFTNFGO
CWLKWLPEGERAKKPIINLEGIIPCEELAAIILNKFNDAGFLFLATKRGV
KVSUDAFSPNPKRKGIRALEIDGDELIAACHIVSDEKQVLMFLHGMARPPHEKVR
PMGRTARGVSKLNEEDKVVSCQIVTENOSVLVDCQFGKRLSLVDFEFTNKGCV
GVRSLINERNGNVLGAIPTVDHDSILMSSQGAIRINMQDVRVYMGSRSTQGVRLVHL
KEGDALVSMKLSNENDEVLSSSEECSTVSLR"
complement(8391..10808)
/sequence="gyrB_1"
complement(8391..10808)
/sequence="gyrB_1"
/codon_start=1
/transl_table=11
/product="DNA gyrase subunit B"
/protein_id="BAA98485.1"
/db_xref="GI:8978649"
/translation="MDPKENYDASAITVLEGLQAVRPGMYIGDTGITGUHLVYE
VVDNSIDEAMAGYCSRIDVRILEDGGIVIVDNGRGTPIEHERESAKQREYSALEV
LTVLHAGKFDKDSYKVGSLHGVSCVNALSEKLVAIVFKDKCYOMEFSGRGPVT
PLOYVYSRQCEITVEYDPDKIETCTFDRSILMKRLRELAFNLRGITIVFEDDROV
SEDKVTTFEGGIGSVFVNLQNKESLSEPIYICGTRVGDGGEIEFEALQWNSGS
ELFVSTANNIPTROGTHLTGTSTAIIRVINTYIRAHNLKNNKUALTGEDIREGLTA
VLSVKNPQFQETKOKLGNDSVSAQVQVGEALTIFEEPQIARMIVDKVFVAA
QAREAKKARETLRKSALDSARLKLIDCLEKPEKEMYIVEGDSAGGSAKQGRD
RFOATLPIRKLILNVEKARLOKIFONQIEGTIAALGCGIGADNENLSKLRVRIII
MTDADVGSIIHTLLTFTYRHMTALIENEVYIAOPPLYKVSKKDFRYILSEKEM
SYLLMCTNNESSILFKSTERELRGALESFINVLDSVESFINTELEKAIPESELEMY
KEGIGVPLYIAPATQMGGRYLDSEDEKALQAEETHKFLIIEYKVAVFEDIONQ
LKEYGUDISYLIPOKNEIIVIGNEDSPSCNYCYTLEEVINYLKNIIRKRGIEIORYKQ
LGMNADQLWDITMNEQRTLIHVSCLKDAVEADHIFMLMGEEVPPRREFIESHALSI
RINLNDI"
complement(10812..11159)
/sequence="CpJ0276"
complement(10812..11159)
/sequence="CpJ0276"
/codon_start=1
/transl_table=11
/product="Ct191 hypothetical protein"
/protein_id="BAA98486.1"
/db_xref="GI:8978650"
```

```
/translation="MFLKRRKRGSGVQNKRTASPIKHAHYLHNYLQELQK IMAARP
HDAIDAMNOVFRDKYKMSQAIGFRDHIILLVKVYNSSLYALLKOTPDNDLMSLYQVA
SHVOIREIQFLG"
alignment_scores:
  Quality: 883.00      Length: 942
  Ratio: 1.541        Gaps: 31
  Percent Similarity: 60.828  Percent Identity: 30.573
alignment_block:
  US-09-677-752-4 x AP002546/rev ..
  Align seg 1/1 to reverse of: AP002546 from: 1 to: 300650
  12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlycInAspProLe 28
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207773 TGTTCACACTGTTTTCGTCACACTGCTGAAATATATAGGCCCTCTGATAG 207724
  28 u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValValC 43
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207723 CTTTGACGGGAAGTACTAACAACAGCAGCCTTACTCTCTAAAATACGACTA 207674
  43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207673 CTGGAATAGACTATCTCTCAGCAGGAGATATACTCTGCAAAACCTTGGG 207624
  55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207623 GATTTCGCGACTTTTACGAAAGGGTGTCTTCTGACACTACGGAATCTTT 207574
  68 uTYrValLeuGlyAsnSerTrpCysTrpPheValSerLysLeuHisIle 85
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207573 AAGCTTTCGCGGTAAGGGGTACTCA.....CTTCTCTTTTAAATATTA 207530
  85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleGln 101
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207529 AGCTACTGCTGAAGGC.....GCAGCACCTTCTCTGTACA 207495
  102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207494 ACTGATAAAATCTGCTCGCTAACAGGATTTTCGAGTCTTACTTTCTTAGC 207445
  116 rSerProSerIleIle.....HisGlnL 124
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207444 GGCCCCATCATCGGTATATACACACCCCTCAGGAAAGGTCAGTTAAAT 207395
  124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207394 GTGAGGGGATCTTACATTTGATAACAATGCACTATTTTATAACAA 207345
  141 AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207344 GATTACTGTAGGAAATATGCGGAGCCATTTCTACCAAGAATCTTCTTT 207295
  157 uGlnHisAsnTyrLeuPheThrAlaPheGluAsnSerSerLysGly. 173
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207294 GAAAAACAGCAGCGGATCGATTCTTTTGAAGGAATAAATCGAGCGCA 207245
  174 .....AsnGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207244 CAGGAAAAAGGTCGGCTATTTCGTCTACTGGTACTGTAGATATTACA 207195
  187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207194 AATAATACGGCTCTACCCCTCTCTCGAACAAATATTGCTGAAGTGCAGG 207145
  203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207144 TGGAGCTAATAATACAGCAAGAACTGTACAATTACAGGGAATACGCTCT 207096
  219 roLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
  |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
  207095 ...CTTGTATTTTCTCAAAATAGTGTGACAGCCGCCAGGAATGAGCA 207048
```



```
5884 ACAGCGACATTAGCTGGAGCAATACAGGCTACCTTCCGGAATCCTGAGCG 5835
573 gInserThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5834 TCAAGGACCTTTAGTCTCTAATACGCTTTGGGATCTTTTTCAGACATCC 5785
590 lnAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5784 AAGCGATTCAAGGTGTCATAGAGAGAAGTCTTTGACTCTTTGTCAGAT 5735
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5734 CGAGGCTTGGGTCGGGAGTCGCGCAAT...TTCTTAGATAAAGATAA 5688
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5687 GAAAGGGGAAAA...CGCAATACCGTCATAAATCTGGTGATATGCTA 5641
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaAla 656
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5640 TCGAGGCTGAGCGCAAACTGTTCTCAAACTTAATAGCTTTGCGCTTT 5591
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5590 TGGCAACTCTTTGGTAGCGATAAAGATTCTTAGTCGCTAAATCATAC 5541
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5540 TGATACCTATCCAGGAGCCCTTCATATCCACACATTACAGATGTAGTG 5491
688 ..... 688
5490 GCTTCATAGTTGTCCTTAGATAAACTCTCGCTCTGGAGTCATAAA 5441
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5440 CCCCTCTTTTAGAAGGCGAGCTCGCTTAGAGCCAGCTAGTATGATCT 5391
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerTrpHis 721
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5390 GAACACAAAGTATACTCGGTAT...CCTGAGGTCAAAGGTTCTTGGGGGA 5344
721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5343 ATAATGCTTTTAACATGATGTTGGAGCTTCTTCTCATCTTATCTCGAA 5294
738 GlySerGlyLeuPheSerPheSerIlePheSerLysLeuGlnGlyPh 754
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5293 TACCTGCAATCTTTTGATACCTATGCTCCATACATCAAACTGAATCTGAC' 5244
754 eSerClyThrGlnAspGlyPheGluGluSerSerGlyCyluileArgSerP 771
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5243 CTATATACGTCAGCACAGCTTCTCGGAGAAAGGTACAGAGGAAGATCTT 5194
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5193 TTGATGACAGCAACCTCTCAATTTATCTTTTGCCATATAGGCGTGAAGTTT 5144
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTy 804
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5143 GAGAAGTTCTCTGATGTTAATGACTTTTCTTATGATCTGACTTTATCCTA 5094
804 rIleGlnAspLysArgAspValGluSerGlyProValValLeuLeuL 821
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5093 TGTCTCTGATCTTATCCGCAATGATCCCAATGCACTACAGCACTTGTA 5044
821 ysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAla 837
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5043 TCAGCGGAGCCTCTTGGGAAACTTATGCCAATAACTTAGCACGACAGGCC 4994
838 TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeu...GlnTh 853
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4993 TTGCAAGTGCCTGAGGAGCTACACTACGCTTCTCTCTCTGTTGAAGT 4944
```

```
853 rLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerL 870
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4943 GCTCGGCCAGTTTGTCTTGAAGTTCGTGATCCTCAGCGATTATAATG 4894
870 euAspLeuGlyThrThrTyrArgPhe 878
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4893 TAGATCTTGGGGTAAAGTCCCAATC 4868
seq_name: gb_pat1:A81835
seq_documentation_block: 2787 bp DNA PAT 21-JAN-2000
LOCUS A81835
DEFINITION Sequence 9 from Patent WO98598953.
ACCESSION A81835
VERSION A81835.1 GI:6731868
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 2787)
AUTHORS Madsen,A. and Birkelund,S.
TITLE NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL Patent: WO.9858953-A.9 30-DEC-1998.
FEATURES
Location/Qualifiers
source 1..2787
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 811 a 583 c 598 g 795 t
ORIGIN
```

```
alignment_scores:
Quality: 857.00 Length: 865
Ratio: 1.684 Gaps: 24
Percent Similarity: 58.844 Percent Identity: 30.173
alignment_block:
US-09-677-752-4 x A81835
Align seg 1/1 to: A81835 from: 1 to: 2787
74 SerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGluAl 90
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
352 ACGTTTATAGGTTTCTTCGCTATCTTTTATTCCGTCCTCGAAGTTC 401
90 aLeuPheLysGlyLysAspLeuSerIleGlnAsnPheArgPheLeu 107
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 GATAACTACCGCAAGGAGCGCTAGC..... 429
107 erPheThrAspCysSerSerLysGluSerSerProSerIleIleHisGln 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 .....TGCTCTAGC..... 438
124 LysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysAr 140
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
439 .....GGTAGCTTGAAGTTTGACAAAAATGTCAGTTTGCTCTTCAGCA 482
140 gAsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerL 157
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
483 AAATCTTTCACGAGTAAATGGCGGTGCTATCACGCAAAACTCTTTCAT 532
157 euGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly 173
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 TAACAGGAGCTACAACTGTCAGCTGTGTTTCTGAAATACCTCTCTCAAG 582
174 AsnGlyGlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnVa 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583 AAAGCGGAGGCACTTACAGCTTCGATCCGCTTACCATTACTCGAAACCA 632
189 lSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaI 206
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSer 765
|||||
2393 TGAATTGCATCTTGTCTATGCATCAGGAAGGCTTTTAAAGAACAGCGA 2442
|||||
766 GlyGluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuPr 782
|||||
2443 ACAGAACTCGTGAATTTGGANGTAGCGCTTGTGTAATCTTGCCCTTACC 2492
|||||
782 oileGlyIleThrPheGluLysLysSerGln...LysThrArgThrTyrT 798
|||||
2493 TATCGGATCGGATTTGATAGGAATCAGACTGCCAAGATCAACGTACA 2542
|||||
798 yrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSer 814
|||||
2543 ATCTAACTCTTGGT...TATACTGTGGATCTTGTTCGTAGTAACCCGAC 2589
|||||
815 GlyProValValLeuLysAsnAlaValSerTrpAspAlaPrometal 831
|||||
2590 TGTACGCAACACTCGCAATTAGCGGTGATCTTGGAAACCTTCGGTAC 2639
|||||
831 aasnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaL 848
|||||
2640 CAATTGCGACAGACAGCTTTAGTCCCTCGTGCAGGGAACCATTTTGGT 2689
|||||
848 euHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGly 863
|||||
2690 TTAACCTCAATTTCAAGCCCTTAGCCAAATTTCTTTTGAATTCGTGGG 2739
|||||
864 GlnSerHisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
|||||
2740 TCATCTCGCAATACATGTAGACTTAGGACCAAAATACCAATTC 2784
|||||

```

seq_name: gb_pat1:A81851

```

seq_documentation_block:
LOCUS      A81851          3000 bp      DNA          PAT          21-JAN-2000
DEFINITION Sequence 25 from Patent WO9858953.
ACCESSION  A81851
VERSION     A81851.1   GI:6731876
KEYWORDS    unidentified.
SOURCE      unidentified.
            unclassified.
REFERENCE   1 (bases 1 to 3000)
AUTHORS     Madsen,A. and Birkelund,S.
TITLE       NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE
JOURNAL     Patent: WO 9858953-A 25 30-DEC-1998:
            MADSEN ANNA SORIE (DK); BIRKELUND SVEND (DK)
FEATURES    Location/Qualifiers
             1..3000
                /organism="unidentified"
                /db_xref="taxon:32644"
                /note="unnamed protein product"
             259..>3000
                /protein_id="CAB69200.1"
                /db_xref="GI:6731877"
                /translation="MKSQFSLVLSSTLACFTSCSTVFPAATAENIGPSDFDGSNTNG
                TYTPKNTTGDYLTLDLNLGDSALTKCCFSDTTESLSPACKGYSLFNLNKS
                SBEQAALSVTTDKNLSTGFSSTLFLAEPSSVITPSPGAVKCGDGLFDNNGTILF
                KFCBENGGAISTKNLSKNTSGTSGISFEGNKSSATGKKGAICATGTVDTNNTAPT
                LDNNIAEAGGAINGTGNCTITGNTSLFSENSVATAGNGALSGDADVTISGNOS
                VTFSNGOAVANGAIYAKKLTLAGCGGGISFSNNIVOGTTAGNGCAISILAACECSL
                SAPAGDITFNGNAIVATPQTTKRNISIDIGSTAKITNLRASCHSIFFDYDPTANTAA
                DSTDTLNLKADAGNSTDYSGSIVFSEGEKLSDEAKVADNLSTLKQPVTLTAGNLVL
                KRGVTLTGFTAGSSVIMDAGTTLKASTEEVTLGSLPVDLSLGEKKVVIARA
                ASKNALSDGPILLDDNOGNAYENHDLGKTQDFVQLSALGTATTTDVPVPTVAPT
                HYGQGTGWTGWDVDTASTPKTKTATLAWNTGYPNRPGRQPLVPSNLWGSFSDIOA
                IOGVILERSALTLCSDRGFWAAGVANFLDKDKKERRKVRHKSQGYAIGAAOATCSNL
                ISFACQLPGSKDKDLVAKNHTDITYAGAFYIOHITECSCGFIGCLLDKLPGSWPKPLV
                LEQOLAYSNDLKTGYTAYPVKSGWNNAPNMMGLASSHSYPEYLHCFTDYAPYI
                KNLQTLTIRDSFSEKTEGRSFDLSNLPVIGVKFEKFSDCNDFSVDLTLSYVPD
                LIRNDPKCTTALVISGASWETVYANNLRQALQVRAGSHYAFSPMFEVLQGFVEVRGS

```

```

BASE COUNT      867 a      597 c      658 g      878 t
ORIGIN
alignment_scores:
    Quality: 853.50      Length: 904
    Ratio: 1.555      Gaps: 30
    Percent Similarity: 60.730      Percent Identity: 30.752
alignment_block:
US-09-677-752-4 x A81851
Align seg 1/1 to: A81851 from: 1 to: 3000
12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlyGlnAspProLe 28
|||||
316 TGTTCCTACTGTTTTGCTGCAACTGCTGAAATATATAGCCCTCTGATAG 365
|||||
28 u.....GlyGluThrAlaLeuThrLysAsnProAsnHisValValC 43
|||||
366 CTTTGCACGGAAGTACTAACACAGGCACCTATACTCTTAAAAATACGACTA 415
|||||
43 ys.....ThrPhePheGluAspCysThrMetGluSerLeu... 54
|||||
416 CTGGAATAGACTATATCTCTGACAGGAGATATACTCTGCAAAACCTTGGG 465
|||||
55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
|||||
466 GATTCCGCGAGCTTTAACGAGGGTGTGTTTTCTGCACACTACGGAATCTTT 515
|||||
68 uTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeuHisIleT 85
|||||
516 AAGCTTTGCCGGTAAGGGTACTCA.....CTTCTCTTTTAAATATTA 559
|||||
85 hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleGln 101
|||||
560 AGTCTAGTGTGAAGGC.....GCAGCACCTTCTGTTACA 594
|||||
102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
|||||
595 ACTGATAAAATCTGTCGTAACAGGATTTTCGAGTCTTACTTTCTTAGC 644
|||||
116 rSerProSerIleIle.....HisGlnL 124
|||||
645 GGCCCCATCATCGCTAATCACACCCCTCAGAAAAGGTGCAGTTAAAT 694
|||||
124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
|||||
695 GTGGAGGGGATCTTACATTTTGATAACAATGGAACATATTTTAAACAA 744
|||||
141 AsnHisAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
|||||
745 GATTACTGTGAGGAATAATGGCGAGCCCAITTTACCAAGAATCTTTCTTT 794
|||||
157 uGlnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGly. 173
|||||
795 CAAAACAGCAGCGGATCGATTCTTTTGAAGGGAATAAATCGAGCGCAA 844
|||||
174 .....AsnGlyGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
|||||
845 CAGGAAAAAAGTGGGGCTTTTGTGCTACTGCTGTAGATATTACA 894
|||||
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnCl 203
|||||
895 AATAATACGGCTCTACCCCTCTTCTCGAACCAATATTGCTGAAGCTGCAGG 944
|||||
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsnP 219
|||||
945 TGGACCTATAATACCACAGCAAACTGTACAATTACAGGGAATACGTCT. 993
|||||
219 roLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
|||||

```

```
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTy 804
||||| |||::: :::: ||| :::: |||
2767 GAGAACTCTCTCATGTAATGACTTTCTTCTATGACTGACTTATCCCTA 2816
||||| |||::: :::: ||| :::: |||
804 rleGlnAspLeuLysArgAspValGluSerGlyProValValLeuLeuL 821
||||| |||::: :::: ||| :::: |||
2817 TGTTCCTGATCTTATCCGAATGATCCCAANTGCATCAGCAGCAGTGA 2866
||||| |||::: :::: ||| :::: |||
821 yAsnAlaValSerTTPAspAlaProMetAlaAsnLeuAspSerArgAla 837
||||| |||::: :::: ||| :::: |||
2867 TCAGCGGAGCCTCTGGGAACTTATGCCAATAACTTAGCACCACAGGCC 2916
||||| |||::: :::: ||| :::: |||
838 TyrMetPheArg 841
::: |||
2917 TTGCAAGTGGCT 2928
::: |||

seq_name: gb_baz:CPOMP54

seq_documentation_block:
LOCUS CPOMP54 6030 bp DNA BCT 22-JAN-1999
DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.
ACCESSION AJ001311
VERSION AJ001311.1 GI:3255934
KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane
protein 5.
SOURCE
ORGANISM Chlamydia pneumoniae.
Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 6030)
AUTHORS Knudsen,K.
DIRECT SUBMISSION
TITLE Birkeland,S.
JOURNAL Submitted (29-AUG-1997) Knudsen K., Department of Medical
Microbiology and Immunology, University of Aarhus, The Bartholin
building, University of Aarhus, DK-8000 Aarhus C, DENMARK
REFERENCE 2 (bases 1 to 6030)
AUTHORS Knudsen,K., Madsen,A.S., Mygind,P., Christiansen,G. and
Birkeland,S.
IDENTIFICATION of two novel genes encoding 97- to 99-kilodalton
outer membrane proteins of Chlamydia pneumoniae
JOURNAL Infect. Immun. 67 (1), 375-383 (1999)
MEDLINE 99081766
FEATURES
source
1..6030
location/Qualifiers
/organism="Chlamydia pneumoniae"
/isolate="CDC/CWL-029/VR-1310"
/db_xref="taxon:83558"
/clone="pEX2-18"
/clone="B8F3"
/clone_lib="pbluescript"
/dev_stage="elementary body"
/complement(1..2804)
/gene="omp5"
/complement(<1..2742)
/gene="omp5"
/feature="putative lipoprotein"
/feature="putative lipoprotein"
/transl_table=1
/product="outer membrane protein 5"
/protein_id="CAA04671.1"
/db_xref="GI:3255935"
/db_xref="SPTREMBL:O86163"
/translation="MKSQFSLVLSSTLACFTSCSTVFAAENIGPSDSFDGSTNTG
TYTPKNTTIDYTLTGDIQLNLGDSALTKGCFSDTTESLSFAGKGYLSFLNLIK
SAEGAALSVTTDKNLGFSLSLFLAPSSVITTPSGKAVKCGDLTFDNGTILF
KQDCYEEGALSTNLSKNTSGTISFEGNKSSATGKGAICATGTVDTNNAPT
LFSNNIAEAGGAINSTGCTISLVSFSENSVATAGNGALSGDADVTISQNG
VTFSGADITFNGNAIVATTPTTKRNSIDIGSTAKITNLRAISGHISFIYDPTANTAA
DSTDITLNLNKADAGNSTDYSGLVSFSGEKLSEDEAKVADNLTSTLKQPVTLTAGNLVL
KRGVTLTGKPTQAGSSVINDAGTTTLKASTEEVTLTGLSIPVDSISGEGKVVIAASA
AKSNVLDSPGILLNQNAGNENHDGKTQDFSVQLSALGTATTTDVPVPTVATPT
HYGQGTGWTWDDTASTPTKTKTATLAWNTGTLNPERQGLVPNSLWGSFSDIOA
```

```
557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluAr 573
   ::::::::::::::::::::
975 ACAGCGCATTAGCTTGGACCAATACAGGCTACTTCGGAATCCTGAGCG 926
   ::::::::::::::::::::
573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
925 TCAGGACCTTTAGTTCTCTAATAGCCTTTGGGATCTTTTCAGACATCC 876
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
875 AACGATTCAGGTGTCATAGAGAGAGTCTTTGACTCTTTGTTTCAGAT 826
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisaspSe 623
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
825 CGAGGCTCTCGGCTGCGGAGTGCAGCAAT...TTCCTAGATAAAGATAA 779
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
778 GAAGGGGAAAA...CGCAATACCGCTATAAATCTGGTGGATATGCTA 732
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
640 heGlyIleSerThrHisSerLeuAspHisSerPheCysLeuAlaAla 656
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
731 TCGGAGGTGCGAGCCAAACTTGTCTCAAACTTAATAGCTTTGCCTTT 682
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
681 TGCCAACTCTTTGGTAGCGATAAGATTTCTTACTCGCTAAAAATCATAC 632
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
631 TGATACCTATGACGAGGCCCTTCTATATCCACACATTACAGAATGTAGTG 582
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
688 ..... 688
581 GGTTCATAGTTGTCCTTAGATAAACTTCTCGCTCTTGAGGCATATAA 532
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLe 704
   ::::::::::::::::::::
531 CCCCTCGCTTTTAGAAGGCGAGCTCGCTTATAGCCACGCTCACTAATGATCT 482
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerThrPhis 721
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
481 GAAGCAAGATATACTCGGTAT...CTGAGGTGAAGGTTCTTGGGGGA 435
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
721 erValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
   ::::::::::::::::::::
434 ATATGCTTTTAACATGATGTTGGGAGCTTCTCTCATCTTATCCCGAA 385
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
384 TACCTGCTGTTTGTATGATCTATGCTCCATACATCAAACTGAATCTGAC 335
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
754 eSerGlyThrGlnAspGlyPheGluGluSerSerGlyCluIleArgSerp 771
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
334 CTATATAGCTCAGGACAGCTTCTCGGAGAAGGTACAGAAGGAATCTT 285
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
284 TTGATGACAGCAACTCTCAATTTATCTTTGCTATAGGCGTGAAGTTT 235
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyr 804
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
234 GAGAAGTCTCTGATGTAATGACTTTTCTTATGATCTGACTTATCTCTA 185
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
804 rIleGlnAspLeuLysArgaspValGluSerGlyProValValLeuLeuL 821
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
184 TGTTCCTGATCTTATCCGCAATGATGCCAAATGCACACTGACACTGTAA 135
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
821 ysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAla 837
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
134 TCAGCGGAGCCTCTTGGAAACTTATGCCAATAACTTAGCACGACAGGCC 85
   |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
838 TyrMetPheArg 841
```

```
seq_name: gb_bal:AE001587
84 TTGCAAGTCGT 73
::: |||
```

```
seq_documentation_block:
LOCUS AE001587 16448 bp DNA BCT 01-DEC-2000
DEFINITION Chlamydia pneumoniae section 3 of 103 of the complete genome.
ACCESSION AE001587 AE001363
VERSION AE001587.1 GI:4376271
```

```
KEYWORDS
SOURCE Chlamydia pneumoniae CWL029.
```

```
ORGANISM Chlamydia pneumoniae CWL029
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia
```

```
REFERENCE 1 (bases 1 to 16448)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,
```

```
TITLE Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.
JOURNAL Comparative genomes of Chlamydia pneumoniae and C. trachomatis
```

```
MEDLINE Nat. Genet. 21 (4), 385-389 (1999)
PUBMED 99206606
```

```
REFERENCE 2 (bases 1 to 16448)
AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,
```

```
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
```

```
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
Location/Qualifiers
```

```
1..16448
/organism="Chlamydia pneumoniae CWL029"
```

```
/strain="CWL029"
```

```
/db_xref="taxon:115713"
```

```
172..2957
/feature="Polymorphic Outer Membrane Protein G Family;
```

```
contains frame shift"
```

```
172..705
/feature="pmp_3.1"
```

```
/note="CpN0014"
```

```
615..2957
/feature="pmp_3.2"
```

```
/note="CpN0015"
```

```
3196..4971
/feature="pmp_4.1"
```

```
/note="CpN0016"
```

```
3196..5953
/feature="Polymorphic Outer Membrane Protein G Family;
```

```
contains frame shift"
```

```
4874..5953
/feature="pmp_4.2"
```

```
/note="CpN0017"
```

```
6302..7786
/feature="pmp_5.1"
```

```
/note="CpN0018"
```

```
6302..9139
/feature="Polymorphic Outer Membrane Protein G Family;
```

```
contains frame shift"
```

```
7787..9139
/feature="pmp_5.2"
```

```
/note="CpN0019"
```

```
complement(9380..11467)
```

```
/feature="CpN0020"
```

```
complement(9380..11467)
```

```
/feature="CpN0020"
```

```
/codon_start=1
```

```
/transl_table=11
```

```
/product="Predicted OMP [leader peptide: outer membrane]"
```

```
/protein_id="AAD18173.1"
```

```
/db_xref="GI:4376272"
```

```
/translation="MKRCFLFLASFLVGMSSADALTHQEA VKKNSYL SHFKSVSGIV
TIEDGVLAHNNLRITQANKVYVVENTVGOSKLVAHNVNMYRAKTLVCDLYEYEDT
SLSLTGKTRFVYRPIFLPFPFSPIMPPIINFRGGTGGTGLSGYLSYPSISR
KHFSTTFDLSFFKHGVMGFMNHCQKQVFNKSYAHRLAIDNAEADRYRL
```

4263 GACCAGAAATGCTATAGATCTTCATCGAATGCAAAATTTTAAATCTCC 4312
343 luAlaArgAsnGlyAsp...IleLeuPhePheAspProIleValGlnGlu 358
:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 375
4313 GAGCGACTCGGGAATAAAGTTATTTCTATGATCCTATC 4353
359 SerSerSerLysGluSerProLeuProLeuProSerSerLeuGlnAlaSerValTh 375
4353 4353
375 rSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAlaA 392
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 437
4354ACGAGCTCAGGAGCTACTGATAAGCTCTCTTTGA 4387
392 snArgSer.....ValIlePhe 397
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 438
4388 ATAAAGCTGACGCGAGGATCTGGAATACCTATGAAAGCTACATCGTTTC 4437
398 SerSerGluArgLeuSerGluGluGlu...LysThrProAspAsnLeuTh 413
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 448
4438 TCTGGAGAAACTCTCAGAGAGGAACTTAAGAAACCTGACATCTGAA 4487
413 rSerGlnLeuGlnProIleGlnLeuLysSerGlyArgLeuValLeuL 430
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4537
4488 GTCTACATTTTACACAGCGCTGTAGAGCTTGCTGAGGTCCTTAGTATTGA 4537
430 ysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 446
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4587
4538 AAGATGAGTCACTGTAGTGTCAATACTATACGACAGTTCGAGGATCG 4587
447 LeuLeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAsp...Le 462
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4637
4588 AAGTCGTATGAGTGGAGGACTACTTTTGAAGCAAGCGCTGAGGGGT 4637
462 uLysLeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGluL 478
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4687
4638 CACTCTCAATGGCTACCATTAATAGATTCCTTAGATGGGACAAATA 4687
478 ysSerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLys 491
4688 AAGCTATCATTAAGCGCAGCGCAGCAAGTAGGATGTTGCTTATCAGGG 4737
492 IlePheLeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLe 508
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4738
4738 CCTATCATGCTGTAGATGCTCAGGGGAACCTATTATGAGCATCAATACT 4787
508 uLeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLysGlu. 524
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4788
4788 C.....AGTCACAGCAGGCTTTCTTCTTAAGAGCTTTTCTCACAAAG 4831
525GlnSerHisLeuHisLeuProAspGly.....AsnLeuSer 536
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4832
4832 GAACGATGACTACTACAGATATCCCGATACCCCAATTTCTAAATACAG 4881
537 SerHisPheGlyTyrGlnGlyAspTrpThr.PheSerTrpLysAspSerA 553
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4882
4882 AATCACTATGGGTATCAAGGAACTGGAATAATTTGTTGGTTCGACGATG 4931
553 spGluGlyHisSerLeuIleAla.....AsnTrpThrProLysAsnTyr 567
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4932
4932 CAACCTGCAAAACAAAATGCTACCTTAATCTGGACTTAAACAGGATAC 4981
568 ValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAs 584
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 4982
4982 AAGCGAATCCAGACGCTCAGGACCTTTGCTTCTAATAGCTGTGGGG 5031
584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAlaH 601
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5032
5032 TTCTTTTCGATGCTCCGCTCCATTCAGAGCTCATGAGCGGAGCACAA 5081
601 isGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5082
5082 GTTCGTTATCTCGTCAACAAATTTGCGGTATCAGGAATCCCGGAC... 5128

618 PheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisar 634
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5129
5129 TTTTTCATGAAGATCAGAAAGGAAACCAA...CGTAGTATTCGTCTATC 5175
634 gSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHis 651
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5176
5176 TACCGCGGTTATGCATTAGGAGGAGATCTTCACGGCTTCTGAAAT 5225
651 erPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPhe 667
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5226
5226 TCFTTAATTTTGTCTTTTGTACGCTTTTGGCTACGACAAGGACCATCT 5275
668 IleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGln... 683
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5276
5276 GTGGCTAAGAACCATACCCATGTATATGACAGGGGCAATGAGTTACCGACA 5325
684LeuAlaThrSerLeuMetLysIle..... 691
5326 CTTCCGAGAGCTTAAGACCCCTCGCTAAGATTTTGTACAGAAATTCCTACT 5375
692SerAlaGlnAlaCysTyrAsnGlnSerIleHis 702
5376 CCCTACCTTTTCTTCAATGCTCGGTTTCTTATGCCCACATCCGACAAT 5425
703 GluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGlySerTr 719
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5426
5426 AACATGACCACAAAGTACACTGCTATTCCTCTCTGTTAAG...GGAAGCTG 5472
719 pHisSerValAlaValSerGlyGluValCysAlaSerIleProIleValS 736
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5473
5473 GGAATAATGATCGCTTCGGTATAGAACTGGAGGAGCTATCCCGTAGTTG 5522
736 erAsnGly...SerGlyLeuPheSerPheSerIlePheSerLysLeu 751
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5523
5523 CTTCCAGACGCTCGGCTTGGGTGGATACCCACACGCCATTTCTAAACCTA 5572
752 GlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyGluI 768
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5573
5573 GAGATGATCTATGCACATCAGATGACTTTAAGGAAACCGCAGCAGAGG 5622
768 eaGSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleGlyI 785
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5623
5623 CCGTCTTCCAAAGTGAAGACCTCTTCAATCTAGCGGTTCTCTGTAGGA 5672
785 leThrPheGluLysSerGlnLysThrArgThrTyrTyrPheLeu 801
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5673
5673 TAAATTTGAGAAATTTCCGATAAGTCT...ACGTATGATCTCTCCATA 5719
802 GlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyProValVa 818
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5720
5720 ...GCTTACGTTCCCGATGTGATTGTAATGATCCAGGCTGCAGCAGAAC 5766
818 lLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAsp 835
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5767
5767 TCTTATGTTTCTGGGATCTTGTGCGACATGTTGTTACAGCTTGTGCTA 5816
835 erArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeu...HisArg 850
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5817
5817 GACAAGCTCTTCTTGTACGCTCGAATCATCATGCTTGTCTTCAAC 5866
851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSe 867
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5867
5867 TTTGAAGTTTTCAGTCAGTTTGAAGTCGAGTTGGAGGTTCTTCTCTAG 5916
867 rTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
:||||:||||:||||:||||:||||:||||:||||:||||:||||: 5917
5917 CTATGCTATCATCTTGGAGGAAAGATTCGGATTT 5950

seq_name: gb_ba2:CPN133035

seq_documentation_block:

LOCUS CPN133035 26920 bp DNA BCT 24-FEB-1999

ENRYCCVREQEDYKWEKREAREPRGRGNKILSPREELESSLEQFDHGLKNFSEKLMEL
EGHLLKLOKENTAEVKNKILSDAESRELEIVFEDVKEMPCRIEEIEKTLRWAEPLPLPT
KKAFAKCAQVNSCAEMLEKVKPYCKESLAYVTSKERLVSLDEDLRRAYTECQKRPOG
DSGLSEVRACREGLRIQIFEFTEGLDLVEKELLGVSSRIINTECDVSGVKEAPP
GKKFYAOYDCLYVRVQSRWMTMSERLREGVQACNMKLKACISEEDKVLKEEYWL
REBKNEKELVGLTKIVATOQIOEFOPSDIVESSNEKVSIMDKARLENFREDIS"
12262..12987

gene

/gene="hb6"

CDS

12262..12987

/gene="hb6"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="HB6 protein"

/protein_id="CAB37081.1"

/db_xref="GI:4455897"

/translation="MTAAPATLHVSPPTPEETKEVIPKDSKSRALGITLLVVGILLVV
CGATVLSGVSGLSALVCGLGISTISLVGVFLVGLILLKRLLETLLEAKOIAE
TFABELENYIIGSTKSEKIEGSRISOOGLFNRAITOKILDLLESSLSITSFERDL
QLFDEKIELLSGERLLEFTAANLFKOGROVYLNGLADIRAYMGNPNNTKVAWVIE
KAKAVHEFIVLTTMARELEFF"
13014..14585

gene

/gene="hb7"

CDS

13014..14585

/gene="hb7"

/note="hypothetical"

/codon_start=1

/transl_table=11

/product="HB7 protein"

/protein_id="CAB37082.1"

/db_xref="GI:4455898"

/translation="MYGNERLLERLLYNSVOKSYADRLFSYEKTKMVHDPTLPWEE
DKERCAFAKFLQOKILLDYKSIFWLNENDEINLNPWSGLNVTIRKVPQEV
DSEWNKVLQKEDNYFKLIESSEKTEANKKLLSDLVRIEDAKTKFLKQOE
VETRKDLARYGCTVDPKQDTEAKKVELEASLETPLDSIESLVOCLEDDITYWKE
QDVDLARTQLEEQDTEAKREAEADLSRLERLKKSKTMDRAKHIIENAESITW
WTSQEMDKMARLUKILKEDITSVLPEIDELETLSUEELPKLITRELLTSYKLFKI
CSETLLKMTSRYNVOYEVQNLGFKLOGISORFGKQDDFANLEQVQALOKK
RLRELTONFEIQGFNFKEDEFAAKADLYTRSTAEQKNKNDVPCMELFRYYHEEVNKP
LLELMYNCAEDRYRAKLLKCSLRDLDEKELQKEIKKEFYQKKQORHADRSHRTTYQK
LRIAEELALELKKKI"
14934..17479

gene

/gene="omp7"

CDS

14934..17479

/gene="omp7"

/note="putative"

/codon_start=1

/transl_table=11

/product="outer membrane protein 7"

/protein_id="CAB37083.1"

/db_xref="GI:4455899"

/translation="MKLIPRLLLISLVPTLSMSNLIGAATTEELSASNSFDGTTSTTS
FSSKTSATDCTNIVFKDSVVLENVPKTGETQSTSCFKNDAAAGDLNPLGGGFSFTFS

alignment_scores:

Quality: 852.00

Ratio: 1.635

Percent Similarity: 64.163

Percent Identity: 31.158

alignment_block:

US-09-677-752-4 x CPN133035

Align seg 1/1 to: CPN133035 from: 1 to: 26920

120 lleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136

21233 CTGGTGTGGAGGATATAGGAAATCTTATGAAATCCCTCTAG 21282

136 tserPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleSerAla 153

21283 TGAA.....AATGGCGGAGCCATCAATACGA 21308

153 spAlaPheSerLeuGlnIleAsnTyrLeuPheThrAlaPheGluGlnAsn 169

21309 AGACTTTGTCTTTGACTGGGAGTACCGGTTTGTACGCTTCTTGGCAAT 21358
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSer...Le 185

21359 AGCTCTGCGCAACAGGGGAGCGATCTATGCTTCTGGTGACTCTGTGAT 21408

185 uSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeu 202

21409 TTCTGAGAATGCGAGGAATCTTGAGCTTCGGAAACAACAGTCGCAACAT 21458

202 snGlyGlyAlaIleCysCys...SerAsnLeuIleCysSerGlyAsnVal 217

21459 CAGGAGCGCGATCTCTGCTGAAGGGAACCTTGTGATCTCCAATAACCAA 21508

218 AsnProLeuPhePheThrGly...AsnSerAlaThrAsnGlyGlyAla 233

21509 AAT...ATCTTTTTCGATGGCTGCAAGCAACTACAAATGGCGGAGCTAT 21555

233 eCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSerLeu 250

21556 TGATTCT...AACAAAGCAGGGCGCAACCCAGACCCCTTCTTGCATCTTT 21602

250 laCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGly 266

21603 CAGGAATGAGAGCTTGCATTTTCTGAATAACACAGCAGGAAATAGTGA 21652

267 GlyAlaIleTyrAlaLysHisMetValLeuArgTyr...AsnGlyProva 282

21653 GSTGCGATTATACCAAAAATTTGGTGTATCTCCTCAGGAGGAGGAGT 21702

282 lSerPheIleAsnAsnSerAla.....LysIleGlyGlyAla 295

21703 GTTATTTCCTAACAAACAGCTCGAATGCTACTCTCTAAAGAGAGGCGAA 21752

295 leAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGluGlySer 311

21753 TTGCGATTCTAGATTCTGGAGAGATAGCATTTCTCGAGATCTCGGCAAT 21802

312 ValLeuPheGlnAsnAsnSerGlnArgThr.....SerAspGlnGlyLe 326

21803 ATCATTTTCGAGGGCAATACTACGAGCACTACAGGAAGCTCTCGAGTGT 21852

326 uValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLeuG 343

21853 GACCAGAAATGCTATAGATCTTCGCATCGAATGCAAAATTTTAAATCTCC 21902

343 luAlaArgAsnGlyAsp...IleLeuPhePheAspProIleValGlnGlu 358

21903 GACGAGCTCGGGAAATAAAGTTATTTCTATGATCTCTATC..... 21943

359 SerSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerValTh 375

21943

375 rSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAla 392

21944

392 snArgSer.....ValIlePhe 397

21978 ATAAAGCTGACGAGGATCTGGAAATACCTATGAAGGCTACATCGTTTC 22027

398 SerSerGluArgLeuSerGluGlu...LysThrProAspAsnLeuTh 413

22028 TCTGGAGAGAACTCTCAGAGAGGAACTTAAGAAACCTGACATCTGAA 22077

413 rSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeuValLeu 430

22078 GTCTACATTTACACAGGCTGTAGAGCTTGTGCGAGGTCCTTAGTATGA 22127

430 ysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 446

.....


```
688 LeuMetLysIle.....SerAlaGlnAlaCysTyrAsnGluSerI 701
6960 TATATCTCTCTCTGCAAACTACGGCTGTGTCGTGTGTTATGTTCC 6911
692 .....
6910 TACAGAGATTCTGTTCTCTTTTCAGAAACCTTAGCTACACCCATACGG 6861
701 leHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGly 717
6860 ATACGATCTGAAACCAAGTATACACATATCTACTCTTAA...GGA 6814
718 SerTrpHisSerValAlaValSerGlyGluValCysAlaSerIleProII 734
6813 AGCTGGGAATGATAGTTTCGCTTTAGAAATTCGGTGAAGAGCTCCGAT 6764
734 eValSerAsnGlySerClyLeuPheSerPheSerIlePheSerLysL 751
6763 TTGCTTAGTAGAAGTGCTCTATTGAGCAGTACATGCCCTTCATGAAT 6714
751 euGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyGlu 767
6713 TGCAGTTTCTGTCATGACATCAGCAAGGTTTAAAGAACAGGCAACAGAA 6664
768 IleArgSerPheSerAlaSerPheArgAsnIleSerLeuProIleG 784
6663 GCTCGTGAATTTGGAAGTAGCCGCTTGTGNAATCTTGCCCTTACCTACGG 6614
784 ylleThrPheGluLysSerGln...LysThrArgThrTyrTyrP 800
6613 GATCCGATTGTAGAAAGAAATCAGACTGCCAAGATGCAACGTACAATCTAA 6564
800 heLeuGlyAlaTyrIleGlnAspLeuLysArgAspValGluSerGlyPro 816
6563 CTCTGGT...TATACTGTGATCTTGTCTGTAGTAAACCTTCGGTACGATTT 6467
817 vaIValLeuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLe 833
6516 ACAACACTGCGAATTCAGCGTGATCTTGTGAAACCTTCGGTACGATTT 6467
833 uAspSerArgAlaTyrMetPheArgIleuThrAsnGlnArgAlaLeuHis 849
6466 GCGAAGCAAGCTTTAGTCTCTGCGAGCAACCATTTTGTCTTTAACT 6417
850 ..ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSer 865
6416 CAATTTTGAAGCTTTAGCCAAATTTCTTTTGAATTCGCTGGTCACT 6367
866 HisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
6366 CGCAATTACATGTAGACTTAGGAGCAAAATACCAATTC 6328
seq_name: gb_ba2:AP02545
seq_documentation_block:
LOCUS AP02545 299650 bp DNA BCT 08-DEC-2000
DEFINITION Chlamydia pneumoniae J138 genomic DNA, complete sequence,
section 1/4.
ACCESSION AP02545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795
AB033796 AB033798 AB033799 AB036071 AB036072 AB036073
AB036074 AB036075 AB036076 AB036077 AB036078 AB036079
AB036080 AB036081 AB036082
VERSION AP02545.2 GI:9956082
KEYWORDS
SOURCE Chlamydia pneumoniae J138 (strain:J138) DNA.
ORGANISM Chlamydia pneumoniae J138
REFERENCE 1 (sites)
AUTHORS Shirai.M., Hirakawa.H., Kimoto.M., Tabuchi.M., Kishi.F., Ouchi.K.,
Shiba.T., Ishii.K., Hattori.M., Kihara.S., and Nakazawa.T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
```

Nucleic Acids Res. 28 (12), 2311-2314 (2000)

20330349

2 (bases 1 to 299650)

Shirai.M.

Direct Submission

Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.

Mutsunori Shirai, Yamaguchi University School of Medicine,

Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi

755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,

Tel:81-836-22-2227, Fax:81-836-22-2415)

On Aug 31, 2000 this sequence version replaced gi:6172286

gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318

gi:6172320 gi:6172322 gi:6172324 gi:6172326 gi:6172328 gi:6172330

gi:6172332 gi:6172334 gi:6172336 gi:6172338 gi:6172340 gi:6172342

gi:6172344 gi:6172346 gi:6172348 gi:6172350 gi:6172352 gi:6172354

gi:6172356 gi:6172358 gi:6172360 gi:6172362 gi:6172364 gi:6172366

gi:6172368 gi:6172370 gi:6172372 gi:6172374 gi:6172376 gi:6172378

gi:6172380 gi:6172382 gi:6172384 gi:6172386 gi:6172388 gi:6172390

gi:6172392 gi:6172394 gi:6172396 gi:6172398 gi:6172400 gi:6172402

gi:6172404 gi:6172406 gi:6172408 gi:6172410 gi:6172412 gi:6172414

gi:6172416 gi:6172418 gi:6172420 gi:6172422 gi:6172424 gi:6172426

gi:6172428 gi:6172430 gi:6172432 gi:6172434 gi:6172436 gi:6172438

gi:6172440 gi:6172442 gi:6172444 gi:6172446 gi:6172448 gi:6172450

gi:6172452 gi:6172454 gi:6172456 gi:6172458 gi:6172460 gi:6172462

gi:6172464 gi:6172466 gi:6172468 gi:6172470 gi:6172472 gi:6172474

gi:6172476 gi:6172478 gi:6172480 gi:6172482 gi:6172484 gi:6172486

gi:6172488 gi:6172490 gi:6172492 gi:6172494 gi:6172496 gi:6172498

gi:6172500 gi:6172502 gi:6172504 gi:6172506 gi:6172508 gi:6172510

gi:6172512 gi:6172514 gi:6172516 gi:6172518 gi:6172520 gi:6172522

gi:6172524 gi:6172526 gi:6172528 gi:6172530 gi:6172532 gi:6172534

gi:6172536 gi:6172538 gi:6172540 gi:6172542 gi:6172544 gi:6172546

gi:6172548 gi:6172550 gi:6172552 gi:6172554 gi:6172556 gi:6172558

gi:6172560 gi:6172562 gi:6172564 gi:6172566 gi:6172568 gi:6172570

gi:6172572 gi:6172574 gi:6172576 gi:6172578 gi:6172580 gi:6172582

gi:6172584 gi:6172586 gi:6172588 gi:6172590 gi:6172592 gi:6172594

gi:6172596 gi:6172598 gi:6172600 gi:6172602 gi:6172604 gi:6172606

gi:6172608 gi:6172610 gi:6172612 gi:6172614 gi:6172616 gi:6172618

gi:6172620 gi:6172622 gi:6172624 gi:6172626 gi:6172628 gi:6172630

gi:6172632 gi:6172634 gi:6172636 gi:6172638 gi:6172640 gi:6172642

gi:6172644 gi:6172646 gi:6172648 gi:6172650 gi:6172652 gi:6172654

gi:6172656 gi:6172658 gi:6172660 gi:6172662 gi:6172664 gi:6172666

gi:6172668 gi:6172670 gi:6172672 gi:6172674 gi:6172676 gi:6172678

gi:6172680 gi:6172682 gi:6172684 gi:6172686 gi:6172688 gi:6172690

gi:6172692 gi:6172694 gi:6172696 gi:6172698 gi:6172700 gi:6172702

gi:6172704 gi:6172706 gi:6172708 gi:6172710 gi:6172712 gi:6172714

gi:6172716 gi:6172718 gi:6172720 gi:6172722 gi:6172724 gi:6172726

gi:6172728 gi:6172730 gi:6172732 gi:6172734 gi:6172736 gi:6172738

gi:6172740 gi:6172742 gi:6172744 gi:6172746 gi:6172748 gi:6172750

gi:6172752 gi:6172754 gi:6172756 gi:6172758 gi:6172760 gi:6172762

gi:6172764 gi:6172766 gi:6172768 gi:6172770 gi:6172772 gi:6172774

gi:6172776 gi:6172778 gi:6172780 gi:6172782 gi:6172784 gi:6172786

gi:6172788 gi:6172790 gi:6172792 gi:6172794 gi:6172796 gi:6172798

gi:6172800 gi:6172802 gi:6172804 gi:6172806 gi:6172808 gi:6172810

gi:6172812 gi:6172814 gi:6172816 gi:6172818 gi:6172820 gi:6172822

gi:6172824 gi:6172826 gi:6172828 gi:6172830 gi:6172832 gi:6172834

gi:6172836 gi:6172838 gi:6172840 gi:6172842 gi:6172844 gi:6172846

gi:6172848 gi:6172850 gi:6172852 gi:6172854 gi:6172856 gi:6172858

gi:6172860 gi:6172862 gi:6172864 gi:6172866 gi:6172868 gi:6172870

gi:6172872 gi:6172874 gi:6172876 gi:6172878 gi:6172880 gi:6172882

gi:6172884 gi:6172886 gi:6172888 gi:6172890 gi:6172892 gi:6172894

gi:6172896 gi:6172898 gi:6172900 gi:6172902 gi:6172904 gi:6172906

gi:6172908 gi:6172910 gi:6172912 gi:6172914 gi:6172916 gi:6172918

gi:6172920 gi:6172922 gi:6172924 gi:6172926 gi:6172928 gi:6172930

gi:6172932 gi:6172934 gi:6172936 gi:6172938 gi:6172940 gi:6172942

gi:6172944 gi:6172946 gi:6172948 gi:6172950 gi:6172952 gi:6172954

gi:6172956 gi:6172958 gi:6172960 gi:6172962 gi:6172964 gi:6172966

gi:6172968 gi:6172970 gi:6172972 gi:6172974 gi:6172976 gi:6172978

gi:6172980 gi:6172982 gi:6172984 gi:6172986 gi:6172988 gi:6172990

gi:6172992 gi:6172994 gi:6172996 gi:6172998 gi:6173000 gi:6173002

gi:6173004 gi:6173006 gi:6173008 gi:6173010 gi:6173012 gi:6173014

gi:6173016 gi:6173018 gi:6173020 gi:6173022 gi:6173024 gi:6173026

gi:6173028 gi:6173030 gi:6173032 gi:6173034 gi:6173036 gi:6173038

gi:6173040 gi:6173042 gi:6173044 gi:6173046 gi:6173048 gi:6173050

gi:6173052 gi:6173054 gi:6173056 gi:6173058 gi:6173060 gi:6173062

gi:6173064 gi:6173066 gi:6173068 gi:6173070 gi:6173072 gi:6173074

gi:6173076 gi:6173078 gi:6173080 gi:6173082 gi:6173084 gi:6173086

gi:6173088 gi:6173090 gi:6173092 gi:6173094 gi:6173096 gi:6173098

gi:6173100 gi:6173102 gi:6173104 gi:6173106 gi:6173108 gi:6173110

gi:6173112 gi:6173114 gi:6173116 gi:6173118 gi:6173120 gi:6173122

gi:6173124 gi:6173126 gi:6173128 gi:6173130 gi:6173132 gi:6173134

gi:6173136 gi:6173138 gi:6173140 gi:6173142 gi:6173144 gi:6173146

gi:6173148 gi:6173150 gi:6173152 gi:6173154 gi:6173156 gi:6173158

gi:6173160 gi:6173162 gi:6173164 gi:6173166 gi:6173168 gi:6173170

gi:6173172 gi:6173174 gi:6173176 gi:6173178 gi:6173180 gi:6173182

gi:6173184 gi:6173186 gi:6173188 gi:6173190 gi:6173192 gi:6173194

gi:6173196 gi:6173198 gi:6173200 gi:6173202 gi:6173204 gi:6173206

gi:6173208 gi:6173210 gi:6173212 gi:6173214 gi:6173216 gi:6173218

gi:6173220 gi:6173222 gi:6173224 gi:6173226 gi:6173228 gi:6173230

gi:6173232 gi:6173234 gi:6173236 gi:6173238 gi:6173240 gi:6173242

gi:6173244 gi:6173246 gi:6173248 gi:6173250 gi:6173252 gi:6173254

gi:6173256 gi:6173258 gi:6173260 gi:6173262 gi:6173264 gi:6173266

gi:6173268 gi:6173270 gi:6173272 gi:6173274 gi:6173276 gi:6173278

gi:6173280 gi:6173282 gi:6173284 gi:6173286 gi:6173288 gi:6173290

gi:6173292 gi:6173294 gi:6173296 gi:6173298 gi:6173300 gi:6173302

gi:6173304 gi:6173306 gi:6173308 gi:6173310 gi:6173312 gi:6173314

gi:6173316 gi:6173318 gi:6173320 gi:6173322 gi:6173324 gi:6173326

gi:6173328 gi:6173330 gi:6173332 gi:6173334 gi:6173336 gi:6173338

gi:6173340 gi:6173342 gi:6173344 gi:6173346 gi:6173348 gi:6173350

gi:6173352 gi:6173354 gi:6173356 gi:6173358 gi:6173360 gi:6173362

gi:6173364 gi:6173366 gi:6173368 gi:6173370 gi:6173372 gi:6173374

gi:6173376 gi:6173378 gi:6173380 gi:6173382 gi:6173384 gi:6173386

gi:6173388 gi:6173390 gi:6173392 gi:6173394 gi:6173396 gi:6173398

gi:6173400 gi:6173402 gi:6173404 gi:6173406 gi:6173408 gi:6173410

gi:6173412 gi:6173414 gi:6173416 gi:6173418 gi:6173420 gi:6173422

gi:6173424 gi:6173426 gi:6173428 gi:6173430 gi:6173432 gi:6173434

gi:6173436 gi:6173438 gi:6173440 gi:6173442 gi:6173444 gi:6173446

gi:6173448 gi:6173450 gi:6173452 gi:6173454 gi:6173456 gi:6173458

gi:6173460 gi:6173462 gi:6173464 gi:6173466 gi:6173468 gi:6173470

gi:6173472 gi:6173474 gi:6173476 gi:6173478 gi:6173480 gi:6173482

gi:6173484 gi:6173486 gi:6173488 gi:6173490 gi:6173492 gi:6173494

gi:6173496 gi:6173498 gi:6173500 gi:6173502 gi:6173504 gi:6173506

gi:6173508 gi:6173510 gi:6173512 gi:6173514 gi:6173516 gi:6173518

gi:6173520 gi:6173522 gi:6173524 gi:6173526 gi:6173528 gi:6173530

gi:6173532 gi:6173534 gi:6173536 gi:6173538 gi:6173540 gi:6173542

gi:6173544 gi:6173546 gi:6173548 gi:6173550 gi:6173552 gi:6173554

gi:6173556 gi:6173558 gi:6173560 gi:6173562 gi:6173564 gi:6173566

gi:6173568 gi:6173570 gi:6173572 gi:6173574 gi:6173576 gi:6173578

gi:6173580 gi:6173582 gi:6173584 gi:6173586 gi:6173588 gi:6173590

gi:6173592 gi:6173594 gi:6173596 gi:6173598 gi:6173600 gi:6173602

gi:6173604 gi:6173606 gi:6173608 gi:6173610 gi:6173612 gi:6173614

gi:6173616 gi:6173618 gi:6173620 gi:6173622 gi:6173624 gi:6173626

gi:6173628 gi:6173630 gi:6173632 gi:6173634 gi:6173636 gi:6173638

gi:6173640 gi:6173642 gi:6173644 gi:6173646 gi:6173648 gi:6173650

gi:6173652 gi:6173654 gi:6173656 gi:6173658 gi:6173660 gi:6173662

gi:6173664 gi:6173666 gi:6173668 gi:6173670 gi:6173672 gi:6173674

gi:6173676 gi:6173678 gi:6173680 gi:6173682 gi:6173684 gi:6173686

gi:6173688 gi:6173690 gi:6173692 gi:6173694 gi:6173696 gi:6173698

gi:6173700 gi:6173702 gi:6173704 gi:6173706 gi:6173708 gi:6173710

gi:6173712 gi:6173714 gi:6173716 gi:6173718 gi:6173720 gi:6173722

gi:6173724 gi:6173726 gi:6173728 gi:6173730 gi:6173732 gi:6173734

gi:6173736 gi:6173738 gi:6173740 gi:6173742 gi:6173744 gi:6173746

gi:6173748 gi:6173750 gi:6173752 gi:6173754 gi:6173756 gi:6173758

gi:6173760 gi:6173762 gi:6173764 gi:6173766 gi:6173768 gi:6173770

gi:6173772 gi:6173774 gi:6173776 gi:6173778 gi:6173780 gi:6173782

gi:6173784 gi:6173786 gi:6173788 gi:6173790 gi:6173792 gi:6173794

gi:6173796 gi:6173798 gi:6173800 gi:6173802 gi:6173804 gi:6173806

gi:6173808 gi:6173810 gi:6173812 gi:6173814 gi:6173816 gi:6173818

gi:6173820 gi:6173822 gi:6173824 gi:6173826 gi:6173828 gi:6173830

gi:6173832 gi:6173834 gi:6173836 gi:6173838 gi:6173840 gi:6173842

gi:6173844 gi:6173846 gi:6173848 gi:6173850 gi:6173852 gi:6173854

gi:6173856 gi:6173858 gi:6173860 gi:6173862 gi:6173864 gi:6173866

gi:6173868 gi:6173870 gi:6173872 gi:6173874 gi:6173876 gi:6173878

gi:6173880 gi:6173882 gi:6173884 gi:6173886 gi:6173888 gi:6173890

gi:6173892 gi:6173894 gi:6173896 gi:6173898 gi:6173900 gi:6173902

gi:6173904 gi:6173906 gi:6173908 gi:6173910 gi:6173912 gi:6173914

gi:6173916 gi:6173918 gi:6173920 gi:6173922 gi:6173924 gi:6173926

gi:6173928 gi:6173930 gi:6173932 gi:6173934 gi:6173936 gi:6173938

gi:6173940 gi:6173942 gi:6173944 gi:6173946 gi:6173948 gi:6173950

gi:6173952 gi:6173954 gi:6173956 gi:6173958 gi:6173960 gi:6173962

gi:6173964 gi:6173966 gi:6173968 gi:6173970 gi:6173972 gi:6173974

gi:6173976 gi:6173978 gi:6173980 gi:6173982 gi:6173984 gi:6173986

gi:6173988 gi:6173990 gi:6173992 gi:6173994 gi:6173996 gi:6173998

gi:6174000 gi:6174002 gi:6174004 gi:6174006 gi:6174008 gi:6174010

gi:6174012 gi:6174014 gi:6174016 gi:6174018 gi:6174020 gi:6174022

gi:6174024 gi:6174026 gi:6174028 gi:6174030 gi:6174032 gi:6174034

gi:6174036 gi:6174038 gi:6174040 gi:6174042 gi:6174044 gi:6174046

gi:6174048 gi:6174050 gi:6174052 gi:6174054 gi:6174056 gi:6174058

gi:6174060 gi:6174062 gi:6174064 gi:6174066 gi:6174068 gi:6174070

gi:6174072 gi:6174074 gi:6174076 gi:6174078 gi:6174080 gi:6174082

gi:6174084 gi:6174086 gi:6174088 gi:6174090 gi:6174092 gi:6174094

gi:6174096 gi:6174098 gi:6174100 gi:6174102 gi:6174104 gi:6174106

gi:6174108 gi:6174110 gi:6174112 gi:6174114 gi:6174116 gi:6174118

gi:6174120 gi:6174122 gi:6174124 gi:6174126 gi:6174128 gi:6174130

gi:6174132 gi:6174134 gi:6174136 gi:6174138 gi:6174140 gi:6174142

gi:6174144 gi:6174146 gi:6174148 gi:6174150 gi:6174152 gi:6174154

gi:6174156 gi:6174158 gi:6174160 gi:6174162 gi:6174164 gi:6174166

gi:6174168 gi:6174170 gi:6174172 gi:6174174 gi:6174176 gi:6174178

gi:6174180 gi:6174182 gi:6174184 gi:6174186 gi:6174188 gi:6174190

gi:6174192 gi:6174194 gi:6174196 gi:6174198 gi:6174200 gi:6174202

gi:6174204 gi:6174206 gi:6174208 gi:6174210 gi:6174212 gi:6174214

gi:6174216 gi:6174218 gi:6174220 gi:6174222 gi:6174224 gi:6174226

gi:6174228 gi:6174230 gi:6174232 gi:6174234 gi:6174236 gi:6174238

gi:6174240 gi:6174242 gi:6174244 gi:6174246 gi:6174248 gi:6174250

gi:6174252 gi:6174254 gi:6174256 gi:6174258 gi:6174260 gi:6174262

gi:6174264 gi:6174266 gi:6174268 gi:6174270 gi:6174272 gi:6174274

gi:6174276 gi:6174278 gi:6174280 gi:6174282 gi:6174284 gi:6174286

gi:6174288 gi:6174290 gi:6174292 gi:6174294 gi:6174296 gi:6174298

gi:6174300 gi:6174302 gi:6174304 gi:6174306 gi:6174308 gi:6174310

gi:6174312 gi:6174314 gi:6174316 gi:6174318 gi:6174320 gi:6174322

gi:6174324 gi:6174326 gi:6174328 gi:6174330 gi:6174332 gi:6174334

gi:6174336 gi:6174338 gi:6174340 gi:6174342 gi:6174344 gi:6174346

gi:6174348 gi:6174350 gi:6174352 gi:6174354 gi:6174356 gi:6174358

gi:6174360 gi:6174362 gi:6174364 gi:6174366 gi:6174368 gi:6174370

gi:6174372 gi:6174374 gi:6174376 gi:6174378 gi:6174380 gi:6174382

gi:6174384 gi:6174386 gi:6174388 gi:6174390 gi:6174392 gi:6174394

gi:6174396 gi:6174398 gi:6174400 gi:6174402 gi:6174404 gi:6174406

gi:6174408 gi:6174410 gi:6174412 gi:6174414 gi:6174416 gi:6174418

gi:6174420 gi:6174422 gi:6174424 gi:6174426 gi:6174428 gi:6174430

gi:6174432 gi:6174434 gi:6174436 gi:6174438 gi:6174440 gi:6174442

gi:6174444 gi:6174446 gi:6174448 gi:6174450 gi:6174452 gi:6174454

gi:6174456 gi:6174458 gi:6174460 gi:6174462 gi:6174464 gi:6174466

gi:6174468 gi:6174470 gi:6174472 gi:6174474 gi:6174476 gi:6174478

gi:6174480 gi:6174482 gi:6174484 gi:6174486 gi:6174488 gi


```

2488 AARCAAGGAGCGAAAGTAATTTCAATTAATAAACCAACGAGACATGAA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAlaH 601
2538 CACA...AGCAACACTTACCAATTCGGTTTGTAGCCAAATATCACAGCCA 2584
601 isGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
2585 CTGGTGGG.....GGCTGTGTTTTTTTGTATATA 2613
618 PheTyrAlaHisAspSerGlyLysProIleAsp...AsnTrpHisHi 633
2614 TATGCCAACCATCTGCAGAGGGCTGAGTTAAATGAGTGAATTA 2663
633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspH 650
2664 TATCTCTACGGCGGCTAATTTTACCTTAATTCCTATGTTTCGGCGCATG 2713
650 isSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACGCTTTTAAATC.....AACAAAGACTTA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG1 683
2740 ACCATAAATGCAACCAATTC..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCystyrAsnGluS 700
2761 .....AATTCA 2767
700 erIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCCTCAGACAGCAAGAGATGATTTTATGACGGTACGACGCAATGCC 2817
716 PheGlySerTrpHisSerValAlaValSerGly 726
2818 ATCAATTCACTTACACATATCAATTCCTGGG 2850
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:us-08-302-832-3
seq_documentation_block:
; Sequence 3, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; DATE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berksstresser, Jerry W

```

```

; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4937 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-3

alignment_scores:
Quality: 148.50 Length: 811
Ratio: 0.408 Gaps: 35
Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-302-832-3 ..
Align seg 1/1 to: US-08-302-832-3 from: 1 to: 4937

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG1 89
||||| : : : : : |||||
778 AACGGACAAGTCTTTTAAATCAACCCAAATGGTATCACAAATAGTAAAGA 827
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : |||||
828 CGCAATTATTAAACACTAATGGCTTTACGGCTTCTACGCTAGACATTCTA 877
100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
||||| : : : : : |||||
878 ACCAAACATCAAGCCCGCTAATTTACCTTCAGCAACCAAGATAAA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
||| : : : : : |||
928 GCGCTCGCTGAATTT...GTGAATCAGGTTTAAATTAATCTCGTAAAGA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
||||| : : : : : |||||
975 CGGCAGTGAATCTTATT.....GGTGGCAAG 1003
150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
||||| : : : : : |||||
1004 TGA..... 1008
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
||||| : : : : : |||||
1009 AACGAGGGTGTGATTAGCGTAATGGTGGCAGCATTTCTTTACTCGCAGG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||| : : : : : |||
1059 GCAAAAAATCACCATCAGCGATATAATAAACCAACCACTTACTTACAGCA 1108
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
||||| : : : : : |||||
1109 TTCCCGCGCTGAAATGAAGCG...GTCAATCTGGCGGATATTTTGGC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
||| : : : : : |||
1156 AAA.....GGCGGTACATTAAATGTC.....CG 1178
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
||||| : : : : : |||||
1179 TGCTGCCACT..... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
||||| : : : : : |||||
1189 .....ATTCGAAACCAAGCTAAACTTTCTGCT 1215

```

2818 ATCAATTACCTACACATATCCATTCTGGGC 2850

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-530-198-3

seq_documentation_block:

: Sequence 3, Application US/08530198

: Patent No. 5869085

: GENERAL INFORMATION:

: APPLICANT: BARENKAMP, STEPHEN J

: APPLICANT: ST. GEME III, JOSEPH W

: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS

: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS

: NUMBER OF SEQUENCES: 8

: CORRESPONDENCE ADDRESS:

: ADDRESS: Shoemaker and Mattare, Ltd

: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

: CITY: Arlington

: STATE: Virginia

: COUNTRY: U.S.A.

: ZIP: 22202-0286

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/530,198

: FILING DATE: 13-DEC-1995

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: BERKSTRESSER, JERRY W

: REGISTRATION NUMBER: 22,651

: REFERENCE/DOCKET NUMBER: JWB-1186

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (703) 415-0810

: TELEFAX: (703) 415-0813

: INFORMATION FOR SEQ ID NO: 3:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 4937 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: US-08-530-198-3

alignment_scores:

Quality:	148.50	Length:	811
Ratio:	0.408	Gaps:	35
Percent Similarity:	44.883	Percent Identity:	17.879

alignment_block:

US-09-677-752-4 x US-08-530-198-3

Align seg 1/1 to: US-08-530-198-3 from: 1 to: 4937

```
73  AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG1 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
778  AAGCGACAGTCTTTTATCAACCCCAATGGTATCACAATAGGTAAAGA 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89  uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
828  CCAATTATACACTAATGGCTTTACGGCTTCTACGCTAGACATTCTTA 877
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100  leGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
878  ACGAAACATCAAGCGGGTATTTTCACCTTCGAGCAACCAACAGATAAA 927
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117  SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
928  GCGCTCGTGAATTT...GTGAATCACGGTTTAAATTACTGTCGGTAAAGA 974
```

```
133  nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
975  CGCAGTGTAAATCTTATT.....GGTGGCAAG 1003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150  leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1004  TGAAG..... 1008
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167  GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1009  AACGAGGTGTGATTAGCGTAAATGGTGGCAGCATTTCTTTACTCGCAGG 1058
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180  aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1059  GCAAAATACCATCAGCATGATATAAACCACCATTCATTACAGCA 1108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192  leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1109  TTGCGCGCGCTGAAATGAAGCG...GTCAATCTGGCGATATTTTGGC 1155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209  SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1156  AAA.....GCGGTAACTTAATGTC.....CG 1178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225  nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179  TGCTGCCACT..... 1188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242  erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1189  .....ATTCGAAACCAAGGTAAACTTTCTGCT 1215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259  SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1216  GATTCTGTAGCAAGATAAAGCGCAATATTGTTCTTCGCCCAAGA 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273  smetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1266  GGTCAACGCGAAATTTGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTA 1315
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290  ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1316  AAGCGGCAAGCTGATGATTACAGCGGATAAGTCACATTAATAACAGGT 1365
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301  GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1366  GCAGTTATCGACCTTTCAGGTAAAGAGGGGAGAACTTACCTTGGCGG 1415
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
317  nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1416  TCACGACGCC.....GCGAAGGTAAACCGCATTCATTAG 1453
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
334  LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1454  CAAGAAACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGCCAAA 1503
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347  .....GlyAspIleLeuPhePheAspPr 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1504  CAAAAAGCGGACGCGCTATTGTGTGGCGGATATTGCTTAATTGAC... 1551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354  oLleValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuG 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1552  .....GCCAATATTA 1561
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371  InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1562  ACGCTCAAGTAGTGTGATGATATCGCTAAACCGGTGGT.....TTGTG 1605
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388  GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG1 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1606  GAGACATCGGCGCATTTATTTATCCATTGACAGCAATGCAATGTTAAAC 1655
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404  uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 4937 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-469-880-3

alignment_scores:

Quality: 148.50 Length: 811

Ratio: 0.408 Gaps: 35

Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:

US-09-677-752-4 x US-08-469-880-3 ..

Align seg 1/1 to: US-08-469-880-3 from: 1 to: 4937

```

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGI 89
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
778 AACGGACAAGCTCTTTTAAATCAACCCAAATGGTATCACAAATAGTAAGA 827
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
828 CGCAATATTAAACACTAATAGCTTTACGGCTCTACGCTAGACATTCTA 877
100 LeuGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
878 ACGAACATCAACAGCGCGTAAATTCACCTTCGAGCAACCAAGATATA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
928 CGCGCTCGCTGAAATTT..GTGAATCACGGTTTAAATTAATCTCGGTAAAGA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
975 CGCAGTGTAAATCTTATT.....GCTGGCAAG 1003
150 leSerAlaaspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
..... : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 TGAAG..... 1008
167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....A 180
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1009 AACGAGGCTGATTAGCGTAAATGTGGCAGCATTCCTTACTCGCAGG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1059 GCAAAAATCACCATCAGCATATATAAACCCACCATTAATCTACAGCA 1108
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1109 TTGCCCGCGCTGAAATGAGCG...GTCAATCTGGCGATATTTTGGCC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1156 AAA.....GGCGTAACATTATGTC.....CG 1178
225 nSerAlaThrAsnGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1179 TCCTGCCACT..... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1189 .....ATTGGAACCAACGAGTAACTTCTGCT 1215
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysH 273

```

```

1216 GATTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTTCCGCCCAAGA 1265
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1266 GGGTGAAGCGCAAAATGGCGGTGTAAATTCCTCCGCTCAAAATCAGCAAGCTA 1315
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
290 ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1316 AAGGCGCAAGCTGATGATTACAGGCGATAAGTCACATTAAACACAGGT 1365
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1366 GCAGTTATCAGCCTTTTCAGGTAAGAAGGGGAGAACTTACCTTGGCGG 1415
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1416 TGACGAGCGC.....GCCGAAGGTAAACCGCATTCATTAATTAG 1453
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1454 CAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCATGTATCAGGCAAA 1503
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
347 .....GlyAspIleLeuPhePheAspPr 354
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1504 GAAAAAGCGGACGCGCTATTGTGTGGGCGGATATTGCGTTAATTGAC... 1551
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
354 olleValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuG 371
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1552 .....GGCAATATTA 1561
371 lnAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1562 ACCTCAAGTAGTGTGTATATCGCTAAACCGGTGGT.....TTTGTG 1605
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGI 404
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1606 GACACATCGGGGCAATTTATTCATTTGACAGCAATGCCAATTTGTTAAAC 1655
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1656 AAAAGAGTGGTTCCTACACCGCTGATGATGATTAACAATTTGAAGCCCAAGACC 1705
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1706 CCCTTCGCAATATACCGT.....ATAATGATGATTCGCCNACA 1746
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLe 447
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1747 GGCACCGGTCAAGCAAGCGACCTTAAATAAATAGCGAACTCAAAACAC 1796
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
447 uLeuIle..... 449
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1797 GCTAACCAATAACAATTTTCAAAATTTCTGAAAAACGCGCTGGACAATGA 1846
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
450 ..MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1847 ATATACGGCATCAAGAAACTTACCGTGTAAAGTCAATCAACATCGGA 1896
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
466 ThrLeuSer...IleProLeuHisSer..... 473
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1897 AGCAACTCCCACTTAATTTCTCCATAGTAAAGGTCAGCGTGGCGGCGGT 1946
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
473 ..... 473
1947 TCAGATTGATGGAGATATTACTTTCTAAAGCGGAAATTTTAACCATTTATT 1996
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
474 .....LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1997 CTGCGCGGATGGTGTATGTTTCATATAAATAATATACGCTTGATCAGGTTT 2046
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

878  ACAGAAACATCAGGCGCGTAATTTACCTTCAGCAACCAAGATATA 927
117  SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
928  CGCCTCGCTCAANT...GTGATCAGGTTTAATTAATCTGTCGTAAGA 974
133  nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
975  CGCAGGTGTAATCTTATT...GGTGCAAG 1003
150  leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
1004  TGA... 1008
167  GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln...Al 180
1009  AACGAGGTGTGATTAGCTAATGTTGGCAGCATTTCTTACTCGCAGG 1058
180  aGlnThrPheSerLeuSerArgAsnValSerPro...I 192
1059  GCAAAAAATCACCATCAGCATATAATAAACCAACCATTAATTACAGCA 1108
192  leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
1109  TTCCCGCGCTGAAATGAAGCG...GTCATCTGGCGATATTTTGGC 1155
209  SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
1156  AAA...GGCGGTACATTATGTC...CG 1178
225  nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrs 242
1179  TGCTGCCACT... 1188
242  erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
1189  ...ATTCGAACCAAGGTAACTTCTGCT 1215
259  SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr...AlaLysHI 273
1216  GATCTGTAGCAAGATATAAAGCGGATATTGTTCTTTCCGCCAAGA 1265
273  sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
1266  GGGTGAAGCGAAATGGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTA 1315
290  ys...IleGlyGlyAlaIleAlaIleGlnSerGly 300
1316  AAGCGCGCAAGCTGATGATTACAGCGATATAAGTCACATTAACCAAGGT 1365
301  GlySerLeuSerIleLeuAlaGlyGlySerValLeuPheGlnAsnAs 317
1366  CCACTATCCACCTTCAGTAAAGAGGGGGAGAACTTACCTTGGCGG 1415
317  nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
1416  TGACGAGCGC...GGCGAAGGTAAACCGGCATTCAATTAG 1453
334  luLysAspAlaIleLeuSerSerLeuGluAlaArgAsn... 346
1454  CAAAGAAACCTCTTTAGAAAAAGGCTCAACCATCATGATATCAGGCAAA 1503
347  ...GlyAspIleLeuPhePheAspPr 354
1504  GAAAGGCGGCGCGCTATTGTGGGGCGGATATTGCGTTAATTGAC... 1551
354  oIleValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuG 371
1552  ...GGCAATATTA 1561
371  lnAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387

```

```

1562  ACCTCAAGTAGTGTGTATATCGCTAAAAACCGTGGT...TTTGTG 1605
388  GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGI 404
1606  GAGACATCGGGCATATTATTATCCATTGACAGCAATGCAATTGTTAAAC 1655
404  uGluGlu...LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
1656  AAAAGAGTGTCTGCTAGACCCCTGATGATTAACAATTGAAGCCGAAGACC 1705
419  roIleGluLeuLysSerGlyArgLeuValLeuLysAspArg... 432
1706  CCCTTCGCAATAATACCGGT...ATNAATGATGAATTCCTCAACA 1746
433  ...AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLe 447
1747  GGCACCGGTGAACGACGACCTAAAAAAATAGCGAACTCAAAACAAC 1796
447  uLeuIle... 449
1797  GCTAACCAATAACAATATTTCAAATTTATCTGAAAAACGCGCTGACAATGA 1846
450  ..MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
1847  ATATAACGGCATCAAGAAACTTACCGTTAATAGCTCAATCAACATCGGA 1896
466  ThrLeuSer...IleProLeuHisSer... 473
1897  AGCAACTCCCACTTAATCTCCATAGTAAAGGTCAGCGTGGCGAGCGGT 1946
473  ... 473
1947  TCAGATTGTAGAGATATTACTTTAAAGCGGAAATTTAAACCATTTATT 1996
474  ...LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
1997  CTGGCGGATGGTGTGATTGTTATATAAAATATTACGCTTGATCAGGGTTT 2046
487  LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503
2047  TTAATATTACCCCGCTCCGTAGCTTTTGAAGGTGGAATACAAAGC 2096
503  rGlu...AsnValGluLeuLeuSerLys... 511
2097  ACAGGACGCGGCAATGCTAAATTTGCGCCAGGCGCACTGTAACCATTA 2146
512  ...GluGlnAsnAsnIleProLeu... 518
2147  CAGGAGAGGGAAAGATTTTCAGGGCTAAACAACGTATCTTTAAACGGAACG 2196
519  ...LeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAs 532
2197  GGTAAAGCTCTGAATATCATTTCTCAGTGAATAATTA...AC 2237
532  pGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspTrpThrPheSer 548
2238  CCACATCTTAGTGGCACAATTAACATATCTGGAATATAACAATTAACC 2287
549  ...TrpLysAspSerAspGluGlyHis 556
2288  AAACACTAGAGAAAGAACCCCTGCTATTGCAACACCGCCATGATTCGCAC 2337
557  SerLeuIle...AlaAsnTrpThrProLy 565
2338  TGAAGCTCAGTCTCTTAATCTAGACAGCGGCAAAATTTTACCTTTAT 2387
565  sAsnTyrValProHisProGluArgGlnSerThr... 576
2388  TAATACATTTCAAGCAATAGCAAGGCTTAACAACACAGATAGAAGCT 2437
576  ... 576
2438  CTCGAGGGGTGAATTTTAACGGGTAAATGGCAACATGTCAATCAATCTC 2487

```

```
1189 .....ATTCGAACCAAGGTAAACTTCTGCT 1215
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
1216 GATTCTGTAGCAAGAGATAAAGCGGCAATATTGTTCTTTCCGCCAANGA 1265
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
1266 GGGTGAAGCGGAATTCGGCGGTGTAAATTCGGCTCAAAATCAGCAAGCTA 1315
290 ys.....IleGlyAlaIleAlaIleGlnSerGly 300
1316 AAGCGGCAAGCTGATGATTACAGGCGATAAAGTCACATTAAAAACAGGT 1365
301 GlySerLeuSerIleLeuAlaGlyGlySerValLeuPheGlnAsnAs 317
1366 GCAGTTATCGACCTTCAGGTAAAGAGGGGAGNAACCTTACCTTGGCGG 1415
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
1416 TGACGACGCG.....GGCGAAGGTAAAGACGGCATTCAATTAG 1453
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
1454 CAAGAAAGACCTCTTTAGAAAAGGCTCAACCATCAATGTATCAGGCAAA 1503
347 .....GlyAspIleLeuPhePheAspPr 354
1504 GANAAGCGGACGCGCTATTGTGTGGCGGATATTCGGTTAATTGAC.. 1551
354 oIleValGlnGluSerSerLysGluSerProLeuProSerSerLeuG 371
1552 .....GGCAATATTA 1561
371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValle 387
1562 ACGCTCAAGTAGTGGTGATATCGCTAAAAACCGGTGGT.....TTGTG 1605
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
1606 GAGACATCGGGCAATATTATCCATTGACAGCAATGCCAATGTTAAAC 1655
404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
1656 AAAAGACTGGTTGTAGACCTGATGATGTAACAATTGAAGCCGAAGACC 1705
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
1706 CCCTTCGCAATAATACCGGT.....ATAATGATGAATTCGCCAACA 1746
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLe 447
1747 GGCACCGGTGAGCAGCGCCCTTAAAAAANTAGCGAATCAAAACAC 1796
447 uLeuIle..... 449
1797 GCTAACCAATACAACATTTTCAAAATTTATCTGAAAAACGCGCTGGACAATGA 1846
450 ..MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
1847 ATATAACGGCATCAAGAAACTTACCGCTTAATAGCTCAATCAACATCGGA 1896
466 ThrLeuSer...IleProLeuHisSer..... 473
1897 AGCAACTCCCNCTTAATCTCCATAGTAAGGTGAGCGTGGCGGAGCGGT 1946
473 ..... 473
1947 TCAGATTGATGAGATATTACTTCTAAAGCGGAAATTTAACCAATTTATT 1996
474 .....LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
.....
```



```
600 laHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsp 616
   || |||||
3022 CCACTGGTGG.....GGCTGTGTTTTTTTGAT 3050

617 LeuPheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTrpHi 632
   : : : : : || : : : : : || : : : : :
3051 ATATATGCCAACCACTTCGCACAGGGCTGAGCTTAAATAAGTCAAT 3100

632 sHisArgSerLeuGlyTyrLeuPheGlyLysThrHisSerLeuAspA 649
   : : : : : || : : : : : || : : : : : ||
3101 TAATATCTCTAACGGCGCTAAATTTTACCTTAAATCCCATGTTCCGGCGG 3150

649 sPHisSerPheCysLeuAlaAlaGlyClnLeuLeuGlyLysSerSerAsp 665
   || : : : : : || : : : : : || : : : : : ||
3151 ATGACGCTTTTAAATC.....AACAAAGAC 3176

666 SerPheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAl 682
   || : : : : : || : : : : : || : : : : : ||
3177 TTAACCAATAATCCACCAATTCA..... 3200

682 aGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnG 699
   || : : : : : || : : : : : || : : : : : ||
3201 .....AATT 3204

699 luSerIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGlu 714
   || : : : : : || : : : : : || : : : : : ||
3205 TCAGCTCTACACAGAGAAAGATGATTTTATGACGGGTACGCACGCAAT 3254

715 GlyPheCysSerThrHisSerValAlaValSerGly 726
   : : : : : || : : : : : || : : : : : ||
3255 GCCATCAATTCAACCTACACATATCATCTCTGGC 3290

seq_name: /cgn_6/ptodata/2/lna/5B_COMB.seq:us-08-469-880-6
```

```
seq_documentation_block:
; Sequence 6, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-516 MIS.vg
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-880-6

alignment_scores:
    Quality: 148.50      Length: 812
    Ratio: 0.408        Gaps: 36
    Percent Similarity: 44.828    Percent Identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-469-880-6

Align seg 1/1 to: US-08-469-880-6 from: 1 to: 9323

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysGI 89
||||| : : : : : |||||
1218 AACGGACAACTCTTTTAAATCAACCCAAATGGTATCACAAATAGTAAAGA 1267

89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
||||| : : : : : |||||
1268 CGCAATTATTAACTAATAGCTTACGGCTTCTACGCTAGACATTCTTA 1317

100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
||||| : : : : : |||||
1318 ACGAAACATCAAGCGCGTAATTTTACCTTCGAGCAACCAAGATAAA 1367

117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
: : : : : || : : : : : || : : : : : ||
1368 CGCGCTCGCTCAAAATT...GTGAATCAGCGTTTAAATTACTGTCGTAAGA 1414

133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
||||| : : : : : |||||
1415 CGCAGCTGTAATCTTATT.....GGTCGCAAG 1443

150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
: : : : : || : : : : : || : : : : : ||
1444 TGAAT..... 1448

167 GluGluAsnSerSerLysGlyAsnGlyGlyAlaIleGln.....Al 180
: : : : : || : : : : : || : : : : : ||
1449 AACGAGGTGTGATTACCGTAAATGGTGGCAGCATTTCTTTACTCGCAGG 1498

180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
||| : : : : : ||| : : : : : |||
1499 GCAAAAAATCACCATCAGCATATATAAACCACCACTTACTTACAGCA 1548

192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
||||| : : : : : |||||
1549 TTGCGCGCGCTGAAATGAAGCG...GTCAATCTGGCGCATATTTTGGCC 1595

209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
: : : : : || : : : : : || : : : : : ||
1596 AAA.....GCCGTACATTAAATGTC.....CG 1618

225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
: : : : : || : : : : : || : : : : : ||
1619 TGCTGCCACT..... 1628

242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
||||| : : : : : |||||
1629 .....ATCGAACCAGGTAACTTTCTGCT 1655

259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
: : : : : || : : : : : || : : : : : ||
```



```

; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-557
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9323 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-617-697-6

alignment_scores:
  Quality: 148.50      Length: 812
  Ratio: 0.408        Gaps: 36
Percent Similarity: 44.828 Percent Identity: 18.227

alignment_block:
US-09-677-752-4 x US-08-617-697-6 ..
Align seg 1/1 to: US-08-617-697-6 from: 1 to: 9323

73 AsnSerTyrCysTProPheValSerLysLeuHisIleThrAspProLysG1 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1218 AACGCACAAAGTCCTTTTAAATCAACCCAAATGGTATCACATAGTAAGA 1267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1268 CGCAATTATTAACTATGCGCTTACGGCTTCTACGCTAGACATTCTA 1317
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 leGlnAsnPheArgPheLeuSerPheThrAspCysSerLysGluSer 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1318 ACGAAACATCAAGCGCGTAATTTACCTTCGAGCAACCAAGATAAA 1367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLysArgAsnAs 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1368 CGCGTCGCTGAAAT...GTGAATCAGCGTTTAATTACTGCGTAAAGA 1414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
133 nGlySerMetSerPheCysArgAsnHisAlaGluLysSerGlyAlaI 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1415 CGCAGGTGTAATCTTATT.....GGTGGCAAG 1443
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
150 leSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1444 TGAAG..... 1448
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
167 GluGlnAsnSerSerLysGlyAsnGlyAlaIleGln.....Al 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1449 AACGAGGCGTGATTACGCTAAATGGTGGCAGCATTTCTTACTCGCAGG 1498
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1499 GCAAAATACCATCAGCATATTAATAAACCCCAACCATTTACTACAGCA 1548
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleCysCys 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1549 TTGCGCGCGCTGAAATCAAGCG...GTCAATCTGGCGATATTTTGGC 1595
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1596 AAA.....GGCGGTAAACATTAAATGTC.....CG 1618
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1619 TGCTGCCACT..... 1628
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

1629 .....ATTGAAACCAAGGTAAACTTTCTGCT 1655
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr.....AlaLysHi 273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1656 GATTCTGTAGCAAGATAAAGCGCAATATTGTTCTTCCGCCAAAGA 1705
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
273 smetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1706 GGTGAACGCGAAATTCGCGGTGTAATTTCCGCTCAAAATCAGCAAGCTA 1755
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 ysIleGlyGlyAlaIleAlaIleGlnSer..... 299
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1756 AA...GGCGCAAGCTGATGATAAAGTCCGATAAAGTCACATTAACAA 1802
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
300 GlyGlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAs 316
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1803 GGTGCAGTTATCGACCTTTTCAGGTAAAGAGGGGAGAAACTTACCTTGG 1852
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
316 nAsnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1853 CGGTGACGAGCGC.....GGCAAGGTAAACAAACGCGATTCAAT 1890
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1891 TAGCAAGAAAAACCTCTTTACAAAAAGGCTCAACCATCAATGTATCAGGC 1940
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 .....GlyAspIleLeuPhePheAs 353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1941 AAGAAAAAGCGGACGCGCTATTGTTGGGGCGATATTGCGTTAAATTGA 1990
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 pProIleValGlnGluSerSerLysGluSerProLeuProSerSerL 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1991 C.....GGCAATA 1998
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1999 TTAACGCTCAAGGTAGTGGTGATATCGCTAAACCGGTGGT.....TTT 2042
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2043 GTGAGACATCGGGCATTTATTTATCCATTGACAGCAATGCAATTGTTAA 2092
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2093 AACAAAGAGTGGTGTAGACCTGATGATGATACCAATTGAAGCCGAAG 2142
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
418 lnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2143 ACCCCCTTCGCAATAATACCGGT.....ATAATGATGAATTCCCA 2183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2184 ACAGGCACCGGTGAAGCAAGCGACCTAAAAAAAATAGCGAACTCAAAAC 2233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
446 aLeuLeuIle..... 449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2234 AACGCTACCAATACACTATTTCAAATTATCTGAAAAACGCTGGACAA 2283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 .....MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2284 TGAATATAACGGCATCAAGAAAACTTACCGTTAATAGCTCAATCAACATC 2333
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
465 AlaThrLeuSer...IleProLeuHisSer..... 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2334 GGAAGCAACTCCCACTTAATTTCCATAGTAAAGGTACGCGTGGCGAGG 2383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 ..... 473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2384 CGTTCAGATTGATGAGATATTACTTTCTAAAGCGGGAATTTAACCATTT 2433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2434 ATTCTGGCGGATGGTTGATGTTTCATATAAATATTACCGCTTCATCAGGT 2483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

527 ACGAAACATCAAGCGCGTAATTTCCACCTTGAGCAAAACCAAGGATAAA 576
117 SerProSerIleHISGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
577 GCATCCTGCTGAATC...GTGAATCAGCGTTTAAATACCGTTGGTGAAGA 623
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaI 150
624 CGGTAGCGTAAACCTTATT...GTGGCAAG 652
150 LeSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
653 TGA... 657
167 GluGlnAsnSerSerLysGlyAsnGlyGlyAlaIleGln...Al 180
658 AACGAGCGGTGATTAGCGTAATTCGGGTAGTATTCTTTACTTTCGAGG 707
180 aGlnThrPheSerLeuSerArgAsnValSerPro...I 192
708 GCAAAATCACCATCAGCGATATAATAATCCAAACCATCACATACAGCA 757
192 LeSerPheAlaArgAsnAlaAspLeuAsnGlyGlyAlaIleCysCys 208
758 TTGCTGCACCTGAAACGAGCG...ATCAATCTGGCGATATTTTGGC 804
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPheThrGlyAs 225
805 AAA...GTGGTAACATTAAT... 822
225 nSerAlaThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
823 ...GTCCGCGCTGCCACTATTC 841
242 eGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
842 GCAATAAAGTAAACTTCT...GCC 864
259 SerAsnSerAlaLysGluLysGlyGlyAlaIleTyr...AlaLysHI 273
865 GACTCTGTAGCAAGATAAAGTGAATGATTCCTCTCGCAAGA 914
273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
915 AGGTGAACGCGAATTCGGGTGTAATTCGGCTCAAAATCAGCAAGCCA 964
290 ys...IleGlyGlyAlaIleAlaIleGlnSerGly 300
965 AAGGTGTAAGTTGATGATTACAGCGGATAAAGTTACATTGAAACCGGT 1014
301 GlySerLeuSerIleLeuAlaGlyGluGlySerValLeuPheGlnAsnAs 317
1015 GCATGTATCGACCTTCGGGTAAGAGAGGGGAGAAACTTATCTTGGCG 1064
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
1065 TGACCGCGT...GCGGAGGTAAACCGCATTCATTAAG 1102
334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn... 346
1103 CAAGAAACCACTTTAGAAAAGGCTCAACAATTAATGTGTCAGGTAA 1152
347 ...GlyAspIleLeuPhePheAspPr 354
1153 GAAAGAGTGGCGCGTATGTATGGCGGATATTCGTTAATTTGACG 1202
354 oIleValGlnGluSerSerLysGluSerProLeuProSerSerLeuG 371
1203 CAATATT...AATGCC 1216
371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
1217 AAGTAAAGATATCGCTAAACCTGGTGT...TTTGTG 1251

388 GlnThrSerAla...AsnArgSerValIlePheSe 398
1252 CAGACGTGGCGCATTTATCCATTGATGATAACCGCAATTTGTTAAAC 1301
398 rSerGluArgLeuSerGluGluGlyThrProAspAsnLeuThrSerG 415
1302 AAAAGAATGGCTACTAGAC...CCAGAGAATGTGACTATTG 1339
415 InLeuGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAsp 431
1340 AAGCTCTCTCCGTTCTCGCGTCGAGCTGGT...GCCGAT 1377
432 ArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
1378 AGAATTTCCCACTCGGAGAGGTGATAAA...GTGAC 1412
448 uIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 465
1413 CCTAAAAAATAAACACCTCTTTGACACACTAACCAATACACCATTT 1462
465 laThrLeuSerIleProLeuHis...SerLeuAspThrGluLysSer 479
1463 CAATCTTCTGAAAAGTCCACCGTGTGAACATAACGCAAGGAGAAAA 1512
480 ValThrIleHisAlaProAsnLeuSerIleGlnLys...IlePh 493
1513 CTTACCTCTTAATAGC...TCTATCAGTATAGAAAGGCTCCCACTTAAT 1559
493 eLeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeu... 508
1560 TCTCCACAGTCAGGTGAGGCGGT...CAAGGTGTCAGATTGATA 1603
509 ...LeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLys 523
1604 AAGATATTACTTCTGAAGCGGAAATTTAACCAATTTATCTGCGGATGG 1653
524 GluGlnSerHisLeuHisLeuProAspGly...AsnLeuSe 536
1654 GTTCATGTTTATAAAATATACCTTTGGTAGCGGCTTTTAAACATCAC 1703
536 rSerHisPheGlyTyrGlnGlyAspTyrThrPheSerTrpLysAspSerA 553
1704 AACTAAA...GAGGAGATATCGCTTCGAGACAGTCTGGAC 1744
553 spGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTyrValPro 569
1745 GGAACAACCTAACCATTAACGCCCAAGGAGCCATCAGCTCAGGTAATAGT 1794
570 HisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAsn... 584
1795 AACGGCTTTAGATTTAAACAACGCTCTCTAAACAGCCTTGCGGAAAGCT 1844
585 ThrTyrSerAspMet...GlnAlaValGlnSerMetI 596
1845 GAGCTTTACTGACAGCAGAGGACAGGTAGAGAACTAAGGGTAAATA 1894
596 leAsnThrIleAlaHisGlyGlyAlaTyrLeuPheGlyThrTrpGlySer 612
1895 TCTCAACAATTTGACGGAACGTTAAACATTTCCGGAAC... 1935
613 AlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIleAs 629
1936 ...GTAGATATCTCAATGAAGACCCCAAGT 1964
629 AsnTrpHisHisArgSerLeuGlyTyrLeuPhe...GlyIleSerThrH 645
1965 CAGCTGGTTTTACAGACAAAGACCGCACTACTGGAACGTACCACTT 2014
645 isSerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeuGly 661
2015 TAAATGTT... 2022


```

541 TyrGln...GlyAspTrrPheSerTrpLysAspSerAspGluGlyHi 556
    :   :   :   :   :   :   :   :   :   :   :   :   :
2672 ACAACTATCATGACACAGCAGCGTGGAGTAAGTAAGTCTCAAAATGG 2721
    :   :   :   :   :   :   :   :   :   :   :   :   :
556 sSerLeuLeuAlaAsnTrpThrProLysAsnTyrValProHisProGluA 573
    :   :   :   :   :   :   :   :   :   :   :   :   :
2722 TACAATTAAGGCAACATTAATCTCGCAAAATGTAACAGTGCACGCAACAG 2771
    :   :   :   :   :   :   :   :   :   :   :   :   :
573 rGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMet 589
    :   :   :   :   :   :   :   :   :   :   :   :   :
2772 AAAATCTGTACACAGAGATGCTGTCAATTAATGCAACGCGGCACACA 2821
    :   :   :   :   :   :   :   :   :   :   :   :   :
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLe 606
    :   :   :   :   :   :   :   :   :   :   :   :   :
2822 GTAACATTAAGTACAAAACAGGGGATTAAGAGTGA.....AT 2862
    :   :   :   :   :   :   :   :   :   :   :   :   :
606 uPheGlyThrTrpGly.....SerAlaValSerAsnLeuPheT 619
    :   :   :   :   :   :   :   :   :   :   :   :   :
2863 TGAATCAACTTCGGTAAATGTAATATTAACAGCGCGCAATACACTTA 2912
    :   :   :   :   :   :   :   :   :   :   :   :   :
619 yAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisArgSer 635
    :   :   :   :   :   :   :   :   :   :   :   :   :
2913 AGGTAAGTAATATCACTGGTCAAGATGA..... 2941
    :   :   :   :   :   :   :   :   :   :   :   :   :
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
    :   :   :   :   :   :   :   :   :   :   :   :   :
2942 .....ACAGTAACAGCGGATGCAGAGCGCTT 2967
    :   :   :   :   :   :   :   :   :   :   :   :   :
652 eCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleT 669
    :   :   :   :   :   :   :   :   :   :   :   :   :
2968 GACAACACTACAGCAGCTCAACCATTAAGTGCACAAACAGGCAATGCAATA 3017
    :   :   :   :   :   :   :   :   :   :   :   :   :
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeu... 684
    :   :   :   :   :   :   :   :   :   :   :   :   :
3018 TTACACCAACAAACAGGTGATATCAACCGTAAAGTTGAATCCAGCTCCGGC 3067
    :   :   :   :   :   :   :   :   :   :   :   :   :
685 AlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGluSerIl 701
    :   :   :   :   :   :   :   :   :   :   :   :   :
3068 TCTGTAACTTGTTCGAACCTGGAGCAACTCTGCTGA..... 3106
    :   :   :   :   :   :   :   :   :   :   :   :   :
701 eHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGlys 718
    :   :   :   :   :   :   :   :   :   :   :   :   :
3107 .....GCTAATATTT 3116
    :   :   :   :   :   :   :   :   :   :   :   :   :
718 erTrpHisSerValAlaVal.....SerGlyGluValCysAlaSer 731
    :   :   :   :   :   :   :   :   :   :   :   :   :
3117 CAGTAACACTGTACTATTACTCGGATAGCGGTAAATTAACCTCCACA 3166
    :   :   :   :   :   :   :   :   :   :   :   :   :
732 IleProTleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePh 748
    :   :   :   :   :   :   :   :   :   :   :   :   :
3167 GTAGGTCTACAATTAATGGCACTAATAGTGTACACCACTCAACCCATC 3216
    :   :   :   :   :   :   :   :   :   :   :   :   :
748 eSerLysLeuGlnGly...PheSerGlyThrGlnAspGlyPheGluGly 764
    :   :   :   :   :   :   :   :   :   :   :   :   :
3217 AGCGGATATTGAAGGTACAATTTCTCGTAAATACAGTAAATGTTACAGCAA 3266
    :   :   :   :   :   :   :   :   :   :   :   :   :
764 erSerGlyGluIle 768
    :   :   :   :   :   :   :   :   :   :   :   :   :
3267 GCACCTGGTGATTGA 3280
    :   :   :   :   :   :   :   :   :   :   :   :   :

```

seq_name: /cgn2_6/ptodata/2/ina/5b_COMB.seq:us-08-469-880-7

```

seq_documentation_block:
: Sequence 7, Application US/08459880
: Patent No. 5876733
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J.
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: Of No. 5876733-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,880
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Barksstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4287 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-469-880-7

alignment_scores:
    Quality: 144.00      Length: 788
    Ratio: 0.364        Gaps: 38
    Percent Similarity: 50.254    Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-469-880-7 ..

Align seg 1/1 to: US-08-469-880-7 from: 1 to: 4287

72 GlyAsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLy 88
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 GGAGCGCACCTAC...TGGACGTAAACCACTTTAAATGTTACCTCG..... 1249

88 sGluAlaLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArgP 105
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1250 .....GGTAGTAATTTAACCTCTCCATTGCACG..... 1279

105 heLeuSerPheThrAspCysSerSerLysGluSerSerProSerIleIle 121
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 .....ACAGGAAGTGGCTCAACAGGTCCACGCATA... 1309

122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1310 .....CGCAATGCAGATTAATTAATGTCATAACATTTAATAAGCCACTTT 1353

138 eCysArgAsnHisAlaGluGlySerGlyGly.....AlaIleSerAlaA 153
   |   |||   |||:|||||:|||||:|||||:|||||:|||||:
1354 T.....AATATCCCAAGGCTCAACAGCTAACTTTAGCATCAAGGCAT 1397

153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsn 169
   :   :   :   :   :   :   :   :   :   :   :   :   :
1398 CAATAATGCCCTTTAAGAGTAGTAACCGCTACTACGCATTATTANTAGACAT 1447

```

```

3068 TCTGTAACACTTGTGCAACTGGAGCAACTCTTCTGTGTA..... 3106
701 eHisLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGlyS 718
3107 |||
718 erTrpHisSerValAlaVal.....SerGlyGluValCysAlaSer 731
3117 CAGGTACACTGTTACTATTACTCGGATAGCGGTAAATTAATTAACCTCCACA 3166
732 IleProIleValSerAsnGlySerGlyLeuPheSerPheSerIlePh 748
3167 GTAGGTCTTACAATTAATGGGACTAATAGTGAACCACTCAACGCCAATC 3216
748 eSerLysLeuGlnGly...PheSerGlyThrGlnAspGlyPheGluS 764
3217 AGGCGATATTGAAGGTACAATTTCTGTAATACAGTAATAATGTTACAGCAA 3266
764 erSerGlyGluIle 768
3267 GCACGTGGTGATTA 3280
```

seq_name: /cgn2_6/ptodata/2/ina/59a_COMB.seq:us-08-728-470-7

```

seq documentation_block:
; Sequence 7, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

```

US-08-728-470-7

alignment_scores:
  Quality: 144.00      Length: 788
  Ratio: 0.364        Gaps: 38
  Percent Similarity: 50.254  Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-728-470-7

Align seg 1/1 to: US-08-728-470-7 from: 1 to: 4287

72 GlyAsnSerTyrCysTTPpHeValSerLysLeuHisIleThrAspProLy 88
|||
1208 GGACGCACTTAC...TGAACGTAAACCACTTTAAATGTACTCTCG.... 1249
88 sGluAlaLeuPheLysGlyLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||
1250 .....GGTACTAAATTTAACTCTCCATTGCACAGC..... 1279
105 heLeuSerPheThrAspCysSerSerLysGluSerSerProSerIleIle 121
|||
1280 .....ACAGGAAGTGGCTCAACAGGTCACAGCAT... 1309
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||
1310 .....CGCAATGCAGAAATTAATGGCATTAACATTTAATAAGCCACTTT 1353
138 eCysArgAsnHisAlaGluGlySerGlyGly.....AlaIleSerAla 153
|||
1354 T.....AATATCCACACAGGCTCAACAGCTAACTTTAGCATCAAGCAT 1397
153 sPAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheGluGluAsn 169
|||
1398 CAATAATGCCCTTTAAGAGTAACGCTAACTACGCATTATTTAATGACAGT 1447
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSerLeuSe 186
|||
1448 ATTTCACTCTCAGGGGGGGTAGCGTTAATTTCAAACTTAACGGCTCATC 1497
186 rArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
|||
1498 TAGCAACATACAAACCCCTGCCGTAATTAATAAATCTCAAACTTTAATG 1547
201 euAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
|||
1548 TCTCAGGAGGCTCAACTTTA.....AATCTCAAGGCTGAAGGTTCAACA 1591
218 AsnProLeuPhe.....PheThrGlyAs 225
|||
1592 GAAACCGCTTTTTCATAGAAATGATTTAAACTTAAACGCCACCGGTGG 1641
225 nSerAlaThr..... 228
|||
1642 CAATATAACAATCAGACAGTCGAGGCTACCGATTCAACGCTCAACAAAG 1691
229 .....AsnGlyGlyAlaIleCysCys 235
|||
1692 GTGTCCGAGCCAAAAAACAATACTTTTAAAGGGGTAAATATCACCTTC 1741
236 IleSerAspLeuAsnThrSerGlu...LysGlySerLeuSerLeuAlaCy 251
|||
1742 GGCTCTCAAAAAGCCACACAGAAATCAAGGGCAATGTACCATCAATAA 1791
251 sAsn...GlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGlyG 267
|||
1792 AACACTAACCGCTACTCTCTCGTGGTGGCAATTTTCCGAAACAAATCGC 1841
267 lyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSer 283
|||
1842 CTTTAAATATAGCAGGAATGTTATT.....AATAATGGCAACCTTACC 1885
```

```

STATE: Georgia
COUNTRY: US
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "5'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function= "PolyA_signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type= "terminal"
OTHER INFORMATION: /note= "3'UTR"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product= "Coagulation Factor VIII"
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lakich, D.
AUTHORS: Gitschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 5663060
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 7476
US-08-212-133A-7

```

```

alignment_scores:
Quality: 141.00 Length: 753
Ratio: 0.387 Gaps: 39
Percent Similarity: 48.340 Percent Identity: 21.248

```

```

alignment_block:
US-09-677-752-4 x US-08-212-133A-7

```

```

Align seg 1/1 to: US-08-212-133A-7 from: 1 to: 7493
46 PheGluAspCysThr.....MetGluSerLeuPheProAl 57
|||||
2730 TTCAAAGATTCCACAATTCACAAAATGATATGGAGAGATTGAGCCTCA 2779
|||||
57 aLeu.....CysAlaHisAlaSerGlnAspA 66
|||||
2780 GTTTGAAGAGATAGACAGAGATGCTTAAAGTACAGAGTGTCTCAGTTAGTG 2829
|||||
66 spProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
|||||
2830 ACATGTTGATGCTCTTGGGACAGAGTCATCCT...ACTCCACATGGCTTA 2876
|||||
83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
|||||
2877 TTTTATCAGATGCCCAAGAGCCATCTATCAGGCTATTTCATGATGATCA 2926
|||||
94 .....GluLysGlyAspLeuSerIleGlnAsp 103
|||||
2927 TTCACCAATGCAATAGACAGCAATGAAGCCCATCTAAAGTAGACCAAC 2976
|||||
103 heArgPheLeuSerPheThrAspCysSerSerLysGluSerSerProSer 119
|||||
2977 TCAGG.....CCAGAATCCCATCAGTCAGAGAA 3005
|||||
120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
|||||
3006 ATAGTATTACTCTCTCAGCCGGCTCCAGTTAAGATCCATATAAAAGTTT 3055
|||||
136 tSerPheCysArgAsnHisAlaGluGlySerGlyAlaLysSerAlaA 153
|||||
3056 G.....GAGACAACTATAGAAGTAAAGTGAAGAAACTTGGTTTGC 3096
|||||
153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
|||||
3097 AAGTTTCTAGTTTGCACAGTAATCTAATGACTACACAACTTCTCTCAGAC 3146
|||||
165 .....AlaPheGluGlu..... 168
|||||
3147 AATTGAAAGCACTTTTGAAGACAGACAGATTCTTCAGGATTTCCAGATAT 3196
|||||
169 .....AsnSerSerLysGlyAsnGlyGlyAlaLysGlnAlaGlnT 182
|||||
3197 GCCAGTTCACTCTAGTAGTAATTAAGTACTACTGTCATTTGGTAAGAAAG 3246
|||||
182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
|||||
3247 CATATTCCTTGTGTTGGTCTCATCTACCTTTTAAACGCGAGTGAAGAAAT 3296
|||||
199 AlaAspLeuAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerG1 215
|||||
3297 AGTGAT.....TCCACATATTGGATTCA... 3320
|||||
215 yAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyA 232
|||||
3321 .....ACTTTAATGTATAGTCAAGAAAGTTTACCAAGAGATATA 3360
|||||
232 laIleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
|||||
3361 TATTATCAATAGAGAAATGATAGATTACTACAGAGAGAGAGAGGTTTCATGGA 3410
|||||
247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLys 263
|||||
3411 ATTGCTTTATTGACCAAGATANTACTTTTTCAGAGACAATGTCTCCTT 3460
|||||
263 sGluLysGlyGlyAlaIleTyr.....AlaLysHisMetValLeuA 277
|||||
3461 AATGAACAACAACAACAATATATATCATTCACAACTAATGAAAAACTAC 3510
|||||
277 rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
|||||
3511 AACTGTAGAGGCCCAACATCAATTTGAGAAATAGTACAAACAGACTTTGCAAGAT 3560
|||||

```

```

SEQUENCE CHARACTERISTICS:
  LENGTH: 7493 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: cDNA to mRNA
  HYPOTHETICAL: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
  ORGANISM: Mus musculus
  FEATURE:
    NAME/KEY: repeat_unit
    LOCATION: 1..407
    OTHER INFORMATION: /rpt_type= "terminal"
    OTHER INFORMATION: /note= "5'UTR"
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 7471..7476
    OTHER INFORMATION: /function= "PolyA_signal"
  FEATURE:
    NAME/KEY: repeat_unit
    LOCATION: 7368..7493
    OTHER INFORMATION: /rpt_type= "terminal"
    OTHER INFORMATION: /note= "3'UTR"
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: 408..7367
    OTHER INFORMATION: /product= "Coagulation Factor VIII"
  PUBLICATION INFORMATION:
    AUTHORS: Elder, F.
    AUTHORS: Lakich, D.
    AUTHORS: Gitschier, J.
    TITLE: Sequence of the Murine Factor VIII cDNA.
    Patent No. 5744446
    JOURNAL: Genomics
    VOLUME: 16
    PAGES: 374-379
    DATE: 1993
    RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 7476
  US-08-474-503-5

alignment_scores:
  Quality: 141.00      Length: 753
  Ratio: 0.387         Gaps: 39
  Percent Similarity: 48.340      Percent Identity: 21.248

alignment_block:
  US-09-677-752-4 x US-08-474-503-5

Align seg 1/1 to: US-08-474-503-5 from: 1 to: 7493
46 PheGluAspCysThr.....MetGluSerLeuPheProAl 57
|||||.....|
2730 TTCAAGATTCCACAAATCCCAAAATGATGGAGAAGATTGAGCCTCA 2779
57 aLeu.....CysAlaHisAlaSerGlnAspA 66
|||||.....|
2780 GTTTGACAGATAGCAGAGATGCTTAAAGTACAGAGTCTCTCAGTTAGTG 2829
66 gpProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
|||.....|
2830 ACATGTTGATGCTCTTGGCAGAGATCATCTCT...ACTCCACATGGCTTA 2876
83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
|||||.....|
2877 TTTTATCAGATGGCCAAAGAGACCATCTATGAGGCTATTCATGATGATCA 2926
94 .....GluLysGlyAspLeuSerIleGlnAsnP 103
|||||.....|
2927 TTCACCAATGCAATAGACAGCAATGAGGCCATCTAAAGTGACCCCAAC 2976
103 heArgPheLeuSerPheThrAspCysSerSerSerLysGluSerProSer 119

```

```

2977 TCAGG.....|
120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnGlySerMe 136
|||||.....|
3006 ATAGTATTACTCTCAGCCGGCTCCAGTTAAGATCCAATAAAGTTT 3055
136 tSerPheCysArgAsnHisAlaGluGlySerGlyAlaIleSerAla 153
:.....|
3056 G.....GAGACAACCTATAGAAGTAAAGTGAAGAAACTTGGTTGC 3096
153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
|||||.....|
3097 AAGTTTCTAGTTTCCCAAGTAATCTAATAGTACTACACAATTTCTGCAGAC 3146
165 .....AlaPheGluGlu..... 168
|||||.....|
3147 AATTGAAAGCAACTTTTGAAGAACAGACAGATTCTTCAGGATTTCCAGATAT 3196
169 .....AsnSerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnT 182
|||||.....|
3197 GCCAGTTCACTCTAGTAGTAAATTAAGTACTACTGCATTTTGGTAAGAAAG 3246
182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
|||||.....|
3247 CATATTCCTTGGTGGTCTCATGTACCTTTTAAACGGCAGTGAAGAAAT 3296
199 AlaAspLeuAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGI 215
|||||.....|
3297 AGTGAT.....TCCAACATATTGGATTCA... 3320
215 yAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyA 232
|||||.....|
3321 .....ACTTTAATGTATAGTCAAGAAAGTTTACCACAGAGATAATA 3360
232 laIleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
|||||.....|
3361 TATTATCAATAGAGTAATAGATAGATTACTCAGAGAGAGAGGTTTCATGGA 3410
247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaL 263
|||||.....|
3411 ATTGCTTTATTGACCAAGATAAATCTTTATTCAAAGACAATGTCTCCTT 3460
263 sGluLysGlyGlyAlaIleTyr.....AlaLysHisMetValLeuA 277
|||||.....|
3461 AATGAAACAAACAAACATATATATCATTTCAACAACTAATGAAACACTAC 3510
277 rGlyAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
|||||.....|
3511 ACATGAGAGCCCAACATCAATTTGAGATAGTACACACAGACTTGGCAAGAT 3560
294 AlaIle.....AlaIleGlnSerGlyGlySerLeuSerIle 305
|||||.....|
3561 GCCATATTAAAGTCAATAGTGAGATTCAA.....GAAGTAACAGC 3601
305 eLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrS 322
|||||.....|
3602 TTTGATTCATGATGCAACACTTTTAGGCCAAAAATTTCTACA..... 3641
322 erAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 338
|||||.....|
3642 .....TATTGAGACTAAACCATATATG 3662
339 LeuSer.....SerLeuGluAlaArgAsnGlyAspIleLeuPhe..... 351
|||||.....|
3663 CTAATAGACTACTCTCAACAAAAATAAGACATATTTTCATAGAAAAGA 3712
352 PheAspProIleValGlnGluSerSerSerLysGluSerProLeuProS 368
|||||.....|
3713 TGAAGATCTCTATCCCAAGATGAAGAGAAATACAATCATGCCATTTCCTCA 3762
368 erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
::|

```



```

;
;   AUTHORS: Elder, F.
;   AUTHORS: Lakich, D.
;   AUTHORS: Gitschier, J.
;   TITLE: Sequence of the murine Factor VIII CDNA
;   Patent No. 6180371
;   JOURNAL: Genomics
;   VOLUME: 16
;   PAGES: 374-379
;   DATE: 1993
;   US-09-037-601-5

alignment_scores:
    Quality: 141.00      Length: 753
    Ratio: 0.387         Gaps: 39
    Percent Similarity: 48.340      Percent Identity: 21.248

alignment_block:
US-09-677-752-4 x US-09-037-601-5
Align seg 1/1 to: US-09-037-601-5 from: 1 to: 7493

46  PheGluAspCysThr.....MetGluSerLeuPheProAl 57
|||||  |||  |||||  |||||  |||
2730 TTCAAGATTCCCAAAATTCACAAAATGATATGGAGAAGATTGAGCCTCA 2779

57  aleu.....CysAlaHisAlaSerGlnAspa 66

2780 GTTTGAAGAGATACAGAGATGCTTAAAGTACAGAGTCTCTCAGTTAGT 2829

66  spProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
|||  |||  |||||  |||||  |||
2830 ACATGTCATGCTCTTGGGACAGAGTCATCT...ACTCCACATGGCTTA 2876

83  HisIleThrAspProLysGluAlaLeuPheLys..... 93
|||||  |||  |||||  |||||  |||
2877 TTTTATCAGATGGCCAAAGAGCCATCTATGAGGCTATTTCATGATGATCA 2926

94  .....GluLysGlyAspLeuSerIleGlnAsnP 103

2927 TTCACCAATGCATAGACACCAATGAAGGCCATCTAAAGTGACCCCAAC 2976

103  heArgPheLeuSerPheThrAspCysSerSerLysGluSerSerProSer 119
|||||  |||  |||||  |||||  |||
2977 TCAGG.....CCAGAATCCCATCAGATGAGAAA 3005

120  IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
|||||  |||  |||||  |||||  |||
3006 ATAGTATTACTCTCTCAGCCCGGCTCCAGTTAAGATCCCAATAAAGTTT 3055

136  tSerPheCysArgAsnHisAlaGluGlySerGlyClyAlaIleSerAlaA 153
|||||  |||  |||||  |||||  |||
3056 G.....GAGACAATATAGAGTAAAGTGGGAAGAACTTGGTTGC 3096

153  spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
|||||  |||  |||||  |||||  |||
3097 AAGTTTCTAGTTGGCANGTAACTCTAATGACTACACAACTTCTGTGACAC 3146

165  .....AlaPheGluGlu..... 168
|||||  |||  |||||  |||||  |||
3147 AATTTGAAAGCAACTTTTGAAGAGACAGATCTTCCAGGATTTCCAGATAT 3196

169  .....AsnSerSerLysGlyAsnGlyClyAlaIleGlnAlaGlnT 182
|||||  |||  |||||  |||||  |||
3197 GCCAGTTCACCTAGTAGTAATTAAGTACTACTGCTATTTGGTGAAGAAAG 3246

182  hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
|||||  |||  |||||  |||||  |||
3247 CATATTCCCTTGTGGGTCTCATGTACCTTTAAACGCCGAGTGAAGAAAT 3296

199  AlaAspLeuAsnGlyClyAlaIleCysCysSerAsnLeuIleCysSerG1 215
|||||  |||  |||||  |||||  |||
3297 ACTGAT.....TCCAACTATTGGGATTCA... 3320

```

```

215  yAsnValAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyGlyA 232
|||||  |||  |||||  |||||  |||
3321 .....ACTTTATGTATAGTCAAGAAAGTTTACCAAGAGATATA 3360

232  lalleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
|||  |||  |||||  |||||  |||
3361 TATTATCAATAGAGAATGATAGATTACTACAGAGAGAGAGGTTTTCATGGA 3410

247  LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLy 263
|||||  |||  |||||  |||||  |||
3411 ATTGCTTTATTGACCAAGATAACTATTATTCAAAGACAACTGCTCCTT 3460

263  sGluLysGlyClyAlaIleTyr.....AlaLysHisMetValLeuA 277
|||  |||  |||||  |||||  |||
3461 AATGAAAACAACAACAACATATATCATTCACCAACTAATGAAAAACTAC 3510

277  rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyCly 293
|||  |||  |||||  |||||  |||
3511 ACCTGAGAGCCCAACATCAATTTCAGAATAGTACACAGAGACTTCAAGAT 3560

294  AlaIle.....AlaIleGlnSerGlyGlySerLeuSerIle 305
|||||  |||  |||||  |||||  |||
3561 GCCATATTAAAGGTCAATAGTGAGATTCAA.....GAAGTAACAGC 3601

305  eLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrS 322
|||||  |||  |||||  |||||  |||
3602 TTGTATTCNTGATGGACACTTTTAGCCAAAATTTCTACA..... 3641

322  erAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 338
|||||  |||  |||||  |||||  |||
3642 .....TATTTGAGACTAAACCATATG 3662

339  LeuSer.....SerLeuGluAlaArgAsnGlyAspIleLeuPhe..... 351
|||||  |||  |||||  |||||  |||
3663 CTAATAGAACTAGCTCAACAAAAAATAAGACATATTTTCATAGAAAAA 3712

352  .PheAspProIleValGlnGluSerSerSerLysGluSerProLeuProS 368
|||||  |||  |||||  |||||  |||
3713 TGAGATTCCTATTCCACAGATGAAGAGATACATCATGCCATTTTCCA 3762

368  erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
|||  |||  |||||  |||||  |||
3763 AGATGTTGTTCTTGTCAC.....GAATCTTCAAAAT 3791

385  LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAr 401
|||||  |||  |||||  |||||  |||
3792 TGGTTTAAAAAGACCAATGGAATAATTTCTCTG..... 3824

401  gLeuSerGluGluGlyThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||  |||  |||||  |||||  |||
3825 .AATCTGAGCAGAACAACATAGTCCAAACCAATTAGTATTATTAATG.... 3869

418  lnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaVal 434
|||  |||  |||||  |||||  |||
3870 .....TTTAAAAAATATGTAAAAAATCAAAAGTTTC 3899

435  LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448
|||||  |||  |||||  |||||  |||
3900 TTGTCAGAGAAAAATAAAGTCACAGTAGAACAGGATGATTTTCAACAGAA 3949

448  uIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuA 465
|||  |||  |||||  |||||  |||
3950 CATAGGACTTAAAGACATGGCTTTTCCACATAATATGACATATTTCTTA 3999

465  laThrLeuSerIleProLeuHisSer.....LeuAspThrGluLys 478
|||||  |||  |||||  |||||  |||
4000 CCACCTTGTCT....AACGTACATGAAATGGTAGCCACAATCAAGAAAAA 4046

479  SerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLysIlePh 493
|||||  |||  |||||  |||||  |||
4047 AATATTACGAAGATAGAGAAGCAAGCACTAATTATGAAGAGAAAGTAGT 4096

```

```
57 aLeu.....CysAlaHisAlaSerGlnAspA 66
:::
2780 GTTGNAGAGATACGACAGATGCTTAAAGTACAGAGTGCTCAGTTAGTG 2829
:::
66 spProLeuTyrValLeuGlyAsnSerTyrCysTrpPheValSerLysLeu 82
:::
2830 ACATGTTGATGCTCTTGGGACAGAGTCATCCT...ACTCCACATGGCTTA 2876
:::
83 HisIleThrAspProLysGluAlaLeuPheLys..... 93
:::
2877 TTTTATCATGATGCCAAGAAGACCATCTATCAGGCTATTTCATGATGATCA 2926
:::
94 .....GluLysGlyAspLeuSerIleGlnAsnP 103
:::
2927 TTCACCAATCAATAGACAGCATGAAGCCCATCTAAGTGACCCAC 2976
:::
103 heArgPheLeuSerPheThrAspCysSerSerLysGluSerSerProSer 119
:::
2977 TCAGG.....CCAGAATCCCATCAGTCAGTGA 3005
:::
120 IleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMe 136
:::
3006 ATAGATTACTCTCCAGCCGCGCTCCAGTTAAGATCCCAATAAAGTTT 3055
:::
136 tSerPheCysArgAsnHisAlaGluGlySerGlyGlyAlaIleSerAla 153
:::
3056 G.....GAGACACTAGTAAGTAAGTGAAGAACTGGTTGC 3096
:::
153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThr..... 164
:::
3097 AAGTTTCTAGTTGCCAAGTAATCTAATGACTACAACAATTTCTGTCAGAC 3146
:::
165 .....AlaPheGluGlu..... 168
:::
3147 AATTGAAACCAACTTTTGAAGACAGAGATTCTTCAGAGTTTCAGATAT 3196
:::
169 .....AsnSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnT 182
:::
3197 GCCAGTTCACTCTAGTAGTAATTAAGTACTACTGCTATTTGGTAAGAA 3246
:::
182 hrPheSerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArg 198
:::
3247 CATATTCCTCTGTTGGGTCTCATGACTACTTAAACGCGAGTGAAGAAAT 3296
:::
199 AlaAspLeuAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerG 215
:::
3297 AGTCAT.....TCCAACATATTGGGATTC... 3320
:::
215 yAsnValAsnProLeuPheThrGlyAsnSerAlaThrAsnGlyGlyA 232
:::
3321 .....ACTTATGTATAGTCAAGAAGTTTACCAGAGATATA 3360
:::
232 laIleCysCysIleSerAspLeuAsnThrSerGluLys.....GlySer 246
:::
3361 TATTATCAATAGAGAATCATAGATTACTCAGAGAGAAGAGGTTTCATCGA 3410
:::
247 LeuSerLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLy 263
:::
3411 ATTGCTTTATTGACCAAGAATAATCTATTATTCAAAGACAAATGCTCCTT 3460
:::
263 sGluLysGlyGlyAlaIleTyr.....AlaLysHisMetValLeuA 277
:::
3461 AATGAACAACAACAACATATATATCATTCACAACAATAATGAACAACTAC 3510
:::
277 rgTyrAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyGly 293
:::
3511 ACACGTGAGAGCCCAACATCAATTCAGATAGTACAAACAGACTTCCAAGAT 3560
:::
294 AlaIle.....AlaIleGlnSerGlyGlySerLeuSerIle 305
:::
3561 GCCATATTAAAGGTCATAGTGAGATTCAA.....GAAGTAACAGC 3601
:::
305 eLeuAlaGlyGluGlySerValLeuPheGlnAsnAsnSerGlnArgThrS 322
:::
3602 TTIGATTTCATGATGGAACACTTTTAGCAGCAAAATTTCTACA..... 3641
:::
322 erAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIle 338
:::
3642 .....TATTTGAGACTAAACCATATG 3662
:::
339 LeuSer.....SerLeuGluAlaArgAsnGlyAspIleLeuPhe..... 351
:::
3663 CTAATAGAACTACCTCACAAAAATAAGACATATTTTCATAGAAAGA 3712
:::
352 .PheAspProIleValGlnGluSerSerSerLysGluSerProLeuProS 368
:::
3713 TGAAGATCCTTATCCACAAGATGAAGAGATACAAATCATGCCATTTTCCA 3762
:::
368 erSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
:::
3763 AGATGTTGTTCTTCTGTC.....GAATCTTCAAT 3791
:::
385 LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAr 401
:::
3792 TGGTTTAAAGACCAATGGAATAATTCCTTG..... 3824
:::
401 gLeuSerGluGluGlyThrProAspAsnLeuThrSerGlnLeuGlnG 418
:::
3825 .AACTCTGACAGACATAGTCCAAAGCAATTAGTATATTATTAATG.... 3869
:::
418 lnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaVal 434
:::
3870 .....TTTAAAAATATGTAAAAAATCAAGTTTC 3899
:::
435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448
:::
3900 TTGCACAGAAAAATAAGTCACAGTAGAAGCAGGATGATTTCACAAGAA 3949
:::
448 uIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 465
:::
3950 CATAGGACTTAAAGACATGCGCTTTCCACATAATATGAGCATATTTCTTA 3999
:::
465 laThrLeuSerIleProLeuHisSer.....LeuAspThrGluLys 478
:::
4000 CCACCTTTGTCT...AAGCTACATGAAATGTTAGGCACAAATCAAGAAAA 4046
:::
479 SerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLysIlePh 493
:::
4047 AATATTGAGGAAGATAGAGAAGGAGCCTAATTGAAGAGAAGTAGT 4096
:::
493 eLeuSerAsn.....SerGlyAspGluAsnPheTyrGluAsnV 506
:::
4097 TTTGCCCCAGGTGCGACGAGCAACTGCTCTAAGAAATTTCTTGAAGACA 4146
:::
506 alGluLeuLeuSerLysGluGlnAsn..... 514
:::
4147 TATTGATAGTACGCTAGGCAAGCAAAATAAGTTTATATGAAGTACATGTA 4196
:::
515 .....AsnIleProLeuThrLeuSerLysGluGlnSerHi 527
:::
4197 CCAGTACTTCAAAACATCATCAATCAATAAACAATTCAACAATAACAGTACA 4246
:::
527 sIleHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
:::
4247 GATTCACATG.....GACATTTCTTTAAAGAGG. 4277
:::
544 spTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeuIleAla 560
:::
4278 .....AAGCACAAGGAAACAAATTCAGAAGGCTTGTA 4310
:::
561 AsnTrpThr.....ProLysAsnTyrValProHisProGluArgG 574
:::
4311 AATAAAACCCAGAGAAATGGTAAAAAATAT...CCAAGCCAGAGAAATAT 4357
:::
574 nSerThrLeu.ValAlaAsnThrLeuTrpAsnThrTyrSerAsp.MetG 590
:::
```

```

312 ValLeuPheGlnAsnSerGlnArgThrSerAspGlnGlyLeuValAr 328
1059 .....CAAATAATCACCACGCGATATAATAA 1087
328 gAsnAlaIleTyrLeu.....GluYAspAlaIle...L 339
1088 CCCAACCACTACTTACAGATTGCGCGCTGAAAATGAAGCGGTCAATC 1137
339 euSerSerLeuGluAlaArgAsnGlyAspIleLeuPhePheAspProile 355
1138 TGGCGATATTTCGCNAAGCGGTACATT..... 1169
356 ValGlnGluSerSerSerLysGluSerProLeuProSerSerLeuGlnAl 372
1169 ..... 1169
372 aserValThrSerProThrProAlaThrAlaSerProLeuValIleGlnT 389
1169 ..... 1169
389 hrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGlu 405
1170 ..AATGTCGGTGCCTCACTATTGCAACCAAGGTAACTTTCGCTCAT 1217
406 GluYThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGluLe 422
1218 TCTGTAAGCAAGAT..... 1232
422 uLysSerGlyArgLeuValLeuLysAspArg..... 432
1233 .AAAAGCGCATATTGTTCTTCGCCCAAGAGCGGTGAAGCGGAATG 1281
433 ..AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeu 448
1282 GCGGTGTAATTTCCGCTCAAAATCAGCAGCTAAAGCGGCAAGCTGATG 1331
449 IleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAl 465
1332 ATTACAGCGGATAAAGTACATATAAAGACAGGTGCAGTTATCGACCTTC 1381
465 a..... 465
1382 AGGTAAGAAGCGGAGAACTTACCTTGGCGGTGACGCGCGCGGAAG 1431
466 ..ThrLeuSerIleProLeu...HisSerLeuAspThrGluLysSerVal 480
1432 GTAAAGGCGCATTCANTTGCNAAGNAACCTTTTAGAAAAGCTCA 1481
481 ThrIleHisAlaPro..... 485
1482 ACCATCAATGTATCAGCAAGAAAGAAAGCGGCGGCTATTGTGTGGG 1531
486 .....AsnLeuSerIleGlnLysIle..... 492
1532 CGATATTGCGTTAATTACGCGCAATATAACGCTCAAGCTAGTGTGATA 1581
493 .....PheLeuSerAsnSerGlyAspGluAsnPheTyr 503
1582 TCCTTAAACCGGTGTTTGTGGAGCGCTGCGGCGATGATTATTATCATC 1631
504 GluAsnValGluLeuLeuSerLysGluGlnAsnAsnIleProLeuLeuTh 520
1632 AAAGCAATCAATTGTCGCGCAAGAG..... 1661
520 rLeuSerLysGluGlnSerHisLeuLeuLeuProAspGlyAsnLeuSerS 537
1662 .....TGGTTGTAGACCGCGAT...AATGTATCTA 1689
537 erHisPheGlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSerAsp 553
1690 TTAATGCAGAAACAGCAGGACGCAATACTTCAGAACGCGATGATAC 1739
554 GluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnTyrValProHI 570
1740 ACGGATCCGGGAATAGTGCAGC...ACCCAAACAGCA..... 1775
570 sProGluArgGlnSerThrLeuValAlaAsnThrLeuTrp..... 583
1776 ...AACAAAGAAAAGACAACATTAACAACACCAACTCTTGAGAGATAC 1821
584 .....AsnThrTyrSerAspMetGlnAla.....Val 592
1822 TAAAAAAGGTACCTTTGTTAAACATCAGTCTAATCAACGATCTATGTC 1871
593 GlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPhe..... 607
1872 AATAGCTCCATTAAAT...TATCCAATGGCAGCTTAACCTCTTGGAGTGA 1918
608 .GlyThrTrpGlySerAlaVal.....SerAsnLeuPheTyrAlaH 621
1919 GGGTCGGAGCGGTGGCGGTTGAGATTAAACAAGATATTACCACCGGTG 1968
621 isAspSer.....SerGlyLysProIleAsp 629
1969 ATGATACCAAGAGGTGCAAACTTAACAATTTACTCAGCGCGTGGTGTAT 2018
630 AsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSe 646
2019 GTTCATAAAATATCTCACTCGGCGCAAGTAACATAAACATTACAGC 2068
646 rLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeu..... 660
2069 TAAACAAGATATCGCCTTTGAGAAAGGAAGCAACCAAGTCATTACAGTC 2118
660 ..... 660
2119 AAGGACTATTACCTCAGGCAATCAAAAGGTTTATAGATTAAATATGTC 2168
661 .....GlyLysSerSerAspSerPheIleThrSerThrGluThrTh 674
2169 TCTCTAAACGGCAGCTGGCAGCGCAGTGCNAATTCACCCTAAAGAACCAA 2218
674 rSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu..... 688
2219 TAAATACGCTATCAACAATAAATTTGAAGGGACTTTAAATATTTCAGGCA 2268
689 ..MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeu 704
2269 AAGTGAACATCTCAATGTTTACCTTAAATAATGAAGTGGATGAT... 2315
705 LysThrLysTyrArgSerPheSerLysGluGlyPheGlySerTyrHis... 720
2316 AAATTCAAAGGAGCAGCTTAC.....TGGAAATT 2344
721 .....SerValAlaValSerGlyValCysAlaSerIleProI 734
2345 AACCTCCTTAAATGTTTCCGAGAGTGGCGAGTT.....AACCTCACTA 2388
734 leValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheSerLys 750
2389 TTGACTCCAGAGGAGGATAGTGCAGGACACACTTACCAGCGCTTATAAT 2438
751 LeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSerGlyG 767
2439 TTAACCGGTATATCATTCACCAAGAC.....ACTACCTTTAA 2476
767 uileArgSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleG 784
2477 TGTGACCAAAATGCAAGACTCACTTT...GACATCAAGGCACCAATAG 2523
784 lyIle 785
2524 GGATA 2528

```

```
433 ..AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeu 448
      :::::::::::::: :::: :::: :::: :::: ::::
1282 CGGTGTAATTCGCTCAAAATCAGCAAGCTAAAGCGGCGAAGCTGATG 1331
      :::::::::::::: :::: :::: :::: :::: ::::
449 IleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLeuAl 465
      :::: :::: :::: :::: :::: :::: :::: ::::
1332 ATTACAGCGATAAAGTCACATTAACAAACAGGTGCAGTTATCGACCTTC 1381
      :::: :::: :::: :::: :::: :::: :::: ::::
465 a..... 465
1382 AGGTAAAGAGGGGAGAAACTTACCTTGGCGGTGACAGCGCGCGAAG 1431
      :::: :::: :::: :::: :::: :::: ::::
466 ..ThrLeuSerIleProLeu...HisSerLeuAspThrGluLysSerVal 480
      :::: :::: :::: :::: :::: :::: ::::
1432 GTAAAGGGCATTCAATTCAGCAAGAAACCTCTTTAGAAAAAGGCTCA 1481
      :::: :::: :::: :::: :::: :::: ::::
481 ThrIleHisAlaPro..... 485
      :::: :::: :::: :::: :::: :::: ::::
1482 ACCATCAATGTATCAGCAAGAAAGGCGGACGCGCTATTGTGTGGG 1531
      :::: :::: :::: :::: :::: :::: ::::
486 .....AsnLeuSerIleGlnLysIle..... 492
      :::: :::: :::: :::: :::: :::: ::::
1532 CGATATTGCGTTAATTGACGCAATATTACGCTCAAGGTAGTGTGATA 1581
      :::: :::: :::: :::: :::: :::: ::::
493 .....PheLeuSerAsnSerGlyAspGluAsnPheTyr 503
      :::: :::: :::: :::: :::: :::: ::::
1582 TCGCTAAMACCGGTGTTTGTGGAGACGTGCGGGCATGATTATTATCATC 1631
      :::: :::: :::: :::: :::: :::: ::::
504 GluAsnValGluLeuLeuSerLysGluGlnAsnAsnIleProLeuLeuTh 520
      :::: :::: :::: :::: :::: :::: ::::
1632 AAAGACAATGCAATTCTTACGCCCAAGAG..... 1661
      :::: :::: :::: :::: :::: :::: ::::
520 rLeuSerLysGluGlnSerHisLeuProAspGlyAsnLeuSerS 537
      :::: :::: :::: :::: :::: :::: ::::
1662 .....TGTTGTTAGACCCGGAT...AATGTAICTA 1689
      :::: :::: :::: :::: :::: :::: ::::
537 erHisPheGlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSerAsp 553
      :::: :::: :::: :::: :::: :::: ::::
1690 TTAATCCAGAAACAGCAGGACGCGCAATACTTCAGAGACCATGAATAC 1739
      :::: :::: :::: :::: :::: :::: ::::
554 GluGlyHisSerLeuIleAsnTrpThrProLysAsnTyrValProIle 570
      :::: :::: :::: :::: :::: :::: ::::
1740 ACGGATCCGGGAATAGTCCAGC...ACCCCAAAACGA..... 1775
      :::: :::: :::: :::: :::: :::: ::::
570 sProGluArgGlnSerThrLeuValAlaAsnThrLeuTrp..... 583
      :::: :::: :::: :::: :::: :::: ::::
1776 ...AACAAGAAAGACCAACATTTACAACACAACTCTTGAGAGTATAC 1821
      :::: :::: :::: :::: :::: :::: ::::
584 .....AsnThrTyrSerAspMetGlnAla.....Val 592
      :::: :::: :::: :::: :::: :::: ::::
1822 TAAAAAAGGTACCTTTGTTAACTACACTGCTTAATCAACGCGCATCTATGTC 1871
      :::: :::: :::: :::: :::: :::: ::::
593 GlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeuPhe..... 607
      :::: :::: :::: :::: :::: :::: ::::
1872 AATAGCTCCATTAAT...TTATCCAATGCCAGCTTAACTCTTTGGAGTGA 1918
      :::: :::: :::: :::: :::: :::: ::::
608 ..GlyThrTrpGlySerAlaVal.....SerAsnLeuPheTyrAlaH 621
      :::: :::: :::: :::: :::: :::: ::::
1919 GGTGCGGACGGTGGCGGCGTTGAGATTAAACACGATATTACCACCGGTG 1968
      :::: :::: :::: :::: :::: :::: ::::
621 isAspSer.....SerGlyLysProIleAsp 629
      :::: :::: :::: :::: :::: :::: ::::
1969 ATGATACCAGAGGTGCAAACTTAACAATTTACTACGCGCGCTGGTGTGAT 2018
      :::: :::: :::: :::: :::: :::: ::::
630 AsnTrpHisHisArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSe 646
      :::: :::: :::: :::: :::: :::: ::::
2019 GTTCATAAAAAATATCTCACTCGGGGCGCAAGGTAAACATAAACATTACAGC 2068
      :::: :::: :::: :::: :::: :::: ::::
646 rLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeu..... 660
      :::: :::: :::: :::: :::: :::: ::::
2069 TAAACAAGATATCGCTTTTGAGAAAGGAAGCAACCAAGTCATTACAGGTC 2118
      :::: :::: :::: :::: :::: :::: ::::
660 ..... 660
```



```
seq.documentation_block:
: Sequence 1, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,832
: FILING DATE: 16-SEP-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US pct/us93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-404
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0810
: INFORMATION FOR SEQ. ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 516 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-832-1

alignment_scores:
: Quality: 139.00 Length: 819
: Ratio: 0.364 Gaps: 40
: Percent Similarity: 46.642 Percent Identity: 19.902

alignment_block:
US-09-677-752-4 x US-08-302-832-1 ..
Align seg 1/1 to: US-08-302-832-1 from: 1 to: 516

94 GluLysGlyAspLeuSerIleGln...AsnPheArgPheLeuSerPheTh 109
111:|||||
470 GAAAGTCGCGTACCTTAACGCTTAAGCCACTTCCCTATGTTACTACTT 519
109 AspGlySerSerLeuSerLeuSer..... 117
520 TAGGTGTAACATCTATTCCACANCTGTTTAGCAAGCGCTTACAGAGA 569
111:|||||
118 ProSerIleIleHisGlnLysAsnGlyLysLeuSerLeuArgAsnAsnG 134
570 ATGATGTAGTAGAC.....GGCAGAGCAGCTATGCAAGTAGATGG 610
114 ySerMetSerPheCysArgAsnHisAlaGlyLysSerLysGlyAlaIle 151
611 TAATAAACAATTATCCGCAACAGTGTGAC.....GATATCAATTA 651
```

```
151 erAlaAspAlaPheSerLeuGlnHisAsnTyrlleuPheThrAlaPheGlu 167
652 ATTGGAACAATTTTACATCGACCAAAATGAATGGCGACTTTTACAA 701
168 GluAsnSerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPheSe 184
702 GAAAC.....AACAACTCCGCGTATTCAACCGTGTATCATC 739
184 rLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaSpl 201
740 T.....AACCAATCTCCCAATTAAAGGCAATTTTAGATTT 774
201 euAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
775 CTACGACAGACTCTTTTA.....ATC 797
218 AsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaIleCys 234
798 AACCA.....AATGCTATCACAAATAGG 820
234 scYsIleSerAspLeuAsnThrSerGlyLysGlySerLeuSerLeuAlaC 251
821 TAAAGACCGCAATTTTAACTAATGCTTTACGCTTACGCTTACAGACA 870
251 ysAsnGlnGlnThrLeuPheAlaSerAsn.....SerAlaLys 263
871 TTCTACGAAACATCAAGCGCGTATTTTACCTCGACCAACCAAA 920
264 GluLysGlyAlaIleTyrlleAlaLysHis.....MetValLeuArgTy 278
921 GATTAAGCGCTCGCTGAATGTGATACACGCTTAAATCTACGCGTAA 970
278 rAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyAlaI 295
971 AGAGCGCAGTGAATCTTATGTTGCGCAAGTGAATAAACAAGCGGTGGA 1020
295 leAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaGlyGlySer 311
1021 TTACGCTA...AATGCTGACGACTTTCTTACTCGCAGG..... 1058
312 ValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLeuVal 328
1059 .....CAAAAATCACACCATCACCGCATTAATAATAA 1087
328 gAsnAlaIleTyrlleu.....GluLysAspAlaIle...L 339
1088 CCAACCATTTACTTACACAGATTGCCGCGCTGAATAATGAAGCGTCAATC 1137
339 euSerSerLeuGluAlaArgAsnGlyAspIleLeuPhePheAspProIle 355
1138 TGGCGGATATTTTCCCAAGCGGTACATT..... 1169
356 ValGlnGlnSerSerLysGlySerProLeuProSerSerLeuGlnAl 372
1169 ..... 1169
372 aserValThrSerProThrProAlaThrAlaSerProLeuValIleGln 389
1169 ..... 1169
389 hrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluG 405
1170 ...AATGTCGCGTGCACCTATTGCAAAAGGTAATCTTTCGCGAT 1217
406 GluLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGlu 422
1218 TCTGTAACCAAGAT..... 1232
422 ulysSerGlyArgLeuValLeuLysAspArg..... 432
1233 AAAAGCGCAATATTGTTCTTCGCCCAAGAGGCTGAACCGCAATTG 1281
```

```

4358 TACTACTACAGTAGTAACGGCGTTGG.....GACATTTCA 4395
590 nalaValaInserMet.IleaInthIleAlaInSGLyAlaYLeu 606
4396 GACGTGACACATGCGCTTAACCATTAACGTGTCTACACAGGTATC 4445
607 pheGlyThrTrpGlySerAlaValSerAsnLeuPheTYrAlaInASPse 623
4446 ATT..... 4448
623 rSerGlyLysProIleAspAsnTrpHisArgSerLeuGlyTyrLeuP 640
4449 .....AACAGATGAC.....CACAGCAAGCAAAATGAAACATTC. 4484
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIa 656
4485 .....ATTACTAAATCTTCTTATCAGATTTCTGTGATTAAGACACC 4529
657 GlycInLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGlu 673
4530 ACTCAGACA.....ATAGTTCTGACTCACACATTTGTAACATCAGC 4573
673 rThSerTYrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeuMet 690
4574 ATTTCACCAATAGATCTCAAAAGAGTCATTCCAAAATTTTCTC 4623
690 ysIleSerAlaGlnAlaCysTYrAsnGlnSerIleHisGlnLeuLysThr 706
4624 ATGTTCAAGCA.....TCATCTCATTTTATGACTTTAAGACA 4661
707 Lys 707
4662 AAA 4664

seq_name: /cgn2_6/prodata/2/ina/5a_COMB.seq:US-08-038-682-1

seq_documentation_block:
: Sequence 1, Application US/08038682
: Patent No. 5549897
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GENE III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matlare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038,682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: BERKSTRESSER, JERRY W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-293
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 516 base pairs

```

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-038-682-1

alignment_scores:
Quality: 139.00 Length: 819
Ratio: 0.364 Gaps: 40
Percent Similarity: 46.642 Percent Identity: 19.902

alignment_block:
US-09-677-752-4 x US-08-038-682-1 ..

Align seg 1/1 to: US-08-038-682-1 from: 1 to: 5116

94 GluIysGlyAspLeuSerIleGln..AsnPheArgPheLeuSerPheTh 109
111..... 111
470 GAAAGTCGTCACCTTACCGTTAAAGCCACCTTCCGTCATGTACTATCT 519
109 rAspCys.SerSerLysGluSer..... 117
520 TACGTGTACATCTATTCCACAAATCTGTTTAGCAAGCGCCTTACAAGA 569
118 ProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnG 134
570 ATGATGTAGTACAC.....GGCACAGCCACTATGCAAGTAGATGG 610
134 ySerMetSerPheCysArgAsnHisAlaGlyLysGlyAlaIle 151
611 TAATAAAACCATTAATCCGACACACTGTTGAC.....GATATCATTA 651
151 eAlaAspAlaPheSerLeuGlnHisAsnTYrLeuPheThrAlaPheGlu 167
652 ATTGAAACAATTTTACATCGACCAAAATGAATGTCAGTTTTCACAA 701
168 GluAsnSerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPhe 184
702 GAAAC.....AACNACTCCGCCGTATTCAACGCTGTATAC 739
184 rLeuSerArgAsnValSerProIleSerPheAlaArgAsnAlaAspL 201
740 T.....AACCAATCTCCCAATTAAGAGGATTTTAGATT 774
201 euAsnGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
775 CTACGCAACAGCTTTTTA.....ATC 797
218 AsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaIleC 234
798 AACCA.....AATGTTATCACAAATAGG 820
234 scYsIleSerAspLeuAsnThrSerGluLysGlySerLeuSerLeuAl 251
821 TAAAGACGCAATTAATTAACACTAATGCTTACCGCTTACGCTAGACA 870
251 ysAsnGlnGlnThrLeuPheAlaSerAsn.....SerAlaLys 263
871 TTTCTAACGAAACATCAAGCGCGTAAATTTACCTTCGACCAACCAAA 920
264 GluLysGlyAlaIleTYrAlaLysHis.....MetValLeuArgTY 278
921 GATAAAGCGCTCGCTGAATGTAATGATACGCGTTAATTAATGTCGGTA 970
278 rAsnGlyProValSerPheIleAsnAsnSerAlaLysIleGlyAlaI 295
971 AGACGCGAGTGTAAATCTTATTTGCTGCAAAAGTAAAAACAGCGGTG 1020
295 leaIleGlnSerGlyLysSerLeuSerIleAlaGlyLysSer 311
1021 TTACGCTA...AATGTCGACGACATTTCTTACTCGCAGG..... 1058

```


3900 TTTGTCAGAGAAAATTAAGTCACAGTACAGCATGGATTTTACAAAGAA 3949
448 wllmetclualaglythrserleuylthrserserleuylsleua 465
3950 CATAGAGACTTAAGACATGCGCTTTCCACATTAATGTGACCATTTCTTA 3999
465 lahrleuSerlleProleuHisser.....Leuasprthrluyls 478
4000 CCACCTTGTCT...AAGCTACATGAATGTGAGCACAATCAGAAAAA 4046
479 Serva.....ThrlleHisAlaproasnleuSerlleGlnylsleph 493
4047 AATATTTCAGAGAGATAGACAGAACACTAAATTTGAAGAAAGTGT 4096
493 eleuserasn.....SerGlyAspGluasnPhetylgluasnV 506
4097 TTTGCCCCAGGTGCGACAGACACTGCGCTTACAAATTTCTTGAAGACA 4146
506 aGlnleuLeuSerlySGlnlnasn..... 514
4147 TATTGATCTAGGCACTAGCGCAAAATTAAGTTATATGAGTACATGTA 4196
515AsnllleProleuLeuThrleuSerlySGlnlnSerhi 527
4197 CCAGTACTTCAAAACATCATCATTAACAATTCACAAATACAGTACA 4246
527 sleuHisleuProasprGlyAsnleuSerSerHisphelytyrGlnlyA 544
4247 CATTGACATG.....GACCATTTCTTTTAAACAAGC. 4277
544 sprrprthPheserTrpLysasSeraspGluGlnHisserleuileA 560
4278AAGGCAAGAAACAATTCAGAAAGCGCTGGTA 4310
561 Asntrprthr.....ProlySasnlyValProHisprGluargl 574
4311 AATAAACAAGAGAAATGCTAAAAAACAT...CCAACCAAGCAAGAAAT 4357
574 nserthrleuValAlaAsnThrleuTrpasnThrtyrSerasp_MetG1 590
4358 TACTACTGACAGTACTAAACGCGCTTTGCG.....GACAAATTC 4395
590 naIaValGlnSerMet..IleasnThrIleAlaHisGlyGlyAlaTyrlleu 606
4396 GACTGTCAACTCAATGCTTAAACCATTAACCTGTTCACACAGTATG 4445
607 PhelytyrThrprGlySerAlaValSerAsnleuPhetylAlaHisaspse 623
4446 ATT..... 4448
623 rSerGlyLysProIleasprAsnTrpHisIstargSerleuGlyTyrlleup 640
4449AAACAGATAGAC.....CACAGCAAGAAATGAAAAAGCTTC. 4484
640 heGlyIleSerThrHisSerleuasprHisSerPheCysleuAlaA 656
4485ATTACTAAATCTCTCATCATGATCTCTGTGATTAAGAGCACC 4529
657 GlynleuLeuGlyLysSerSeraspSerPheIlethrSerThrluTh 673
4530 ACTCGACACA.....AATAGTTCCTACACACATTTGAAAAACATCGAC 4573
673 rThrseryrIleAlaThrValGlnAlaGlnleuAlaThrserleuMetL 690
4574 ATTTCGACCATAGATCTCAAAAGAGTCATTCACAAACAATTTTTC 4623
690 yslleSerAlaGlnAlaCysTyranGlnSerIleHisGlnleuLysThr 706
4624 ATGTTCAGACA.....TCATCTACATTTATGACTTTAGACA 4661
707 Lys 707
..

4662 AAA 4664
seq_name: /cgn2_5/prodata/2/lna/5b_COMB.seq:us-09-037-601-5
seq_documentation_block:
Sequence 5, Application US/09037601
Patent No. 6180371
GENERAL INFORMATION:
APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/037,601
FILING DATE: 26-JUN-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 7493 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 1..407
OTHER INFORMATION: /rpt_type="terminal"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7471..7476
OTHER INFORMATION: /function="polya signal"
FEATURE:
NAME/KEY: repeat_unit
LOCATION: 7368..7493
OTHER INFORMATION: /rpt_type="terminal"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 408..7367
OTHER INFORMATION: /product="coagulation factor VIII"
PUBLICATION INFORMATION:

3763 AGANGTGTCTTGTCA.....GAMCTTCAAAAT 3791
 385 LeuValIleGlnThrSerAlaSerSerValIlePheSerSerClar 401
 3792 TGGTTAAAGAACCAATGAAATATATCTTGG..... 3824
 401 gLeuSerGluGluGluYthrProAspAsnLeuThrSerGlnLeuGng 418
 3825 AACCTTCAGCAAGAACATGATCCAAACCAATTTAGTATATTATTAATC.... 3869
 418 LnPProIleGluLeuLysSerGlyArgLeuValIleuLysAspArgAlaVal 434
 3870TTTAAAAATATGTAAAAAATCAAAAGTTTC 3899
 435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaLeuLe 448
 3900 TTGTTCACAGAAAAATAAGTCACATGACATGACATGATTTACAAAGAA 3949
 448 uIleuGluAlaGluYthrSerLeuLysYthrSerSerAspLeuLysLeuA 465
 3950 CATAGAGACTTAAGACATGCGCTTTCCACATATATGACATATTTCTTA 3999
 465 LathrLeuSerIleProLeuHISer.....LeuAspThrGluLys 478
 4000 CCACCTTGTCT...AACGTACATGAAATGTGTAGGCACATCAAGAAAA 4046
 479 SerVal.....ThrIleAlaProAsnLeuSerIleGlnLysIlePh 493
 4047 AATATTCAGCAAGACATGACAAAGACACTAATTCAGAAAGAAATGCT 4096
 493 eLeuSerAsn.....SerGlyAspGluAsnPheYthrGluAsnV 506
 4097 TTTCCCCAGCTCCACAGAACCAACGCTCTTACGAATTTCTTGAAGACA 4146
 506 aIGluLeuLeuSerLysGluGlnAsn..... 514
 4147 TATTGATCTAGGCACTAGGCAAAATATTAAGTTATATGAGTACATGTA 4196
 515AsnIleProLeuLeuThrLeuSerLysGluGlnSerHI 527
 4197 CCAGTACTTCAAAACATCATCATATACAAATTCACAAATACAGTACA 4246
 527 sLeuHISLeuProAspGlyAsnLeuSerSerHISPheGlyYthrGluGlyA 544
 4247 GATTCACATG.....GACCATTTCTTTAAAGCAAGC. 4277
 544 sPTPrThrPheSerTrpLysAspSerAspGluGluYHISerLeuIleA 560
 4278AAGGACAAGGAACAAATTCAGAGGCTTGTA 4310
 561 AsnTrpThr.....ProLysAsnYrValProHISProGluArgI 574
 4311 AATAAACCAAGAAATGCTAAAAAATAT...CCAAGCCAGAAAGATAT 4357
 574 nSerThrLeu.ValAlaAsnThrLeuTrpAsnThrYrSerAsp.MetGI 590
 4358 TACTACTCAACAGTAAAGCGGCTTGG.....GACAAATCA 4395
 590 nAlaValGlnSerMet.IleAsnThrIleAlaHISGlyAlaIleYrLeu 606
 4396 GACGTCACTCAATGCTTAAACCAATAACTGTTCACACAGTGTATC 4445
 607 PheGluYthrTrpGlySerAlaValSerAsnLeuPheYrAlaHISAspSe 623
 4446 ATT..... 4448
 623 rSerGlyLysProIleAspAsnTrpHISHisArgSerLeuGlyYrLeuP 640
 4449AAACAGATAGAC.....CACAGCAAGCAAAATGCAAAAGTTC. 4484
 640 heGlyIleSerThrHISerLeuAspAsnHISerPheCysLeuAlaIleA 656
 4485ATTACTAAATCTTCTTATCAGATCTCTGTGATTAAGACACC 4529

657 GlyIleuLeuGluYLysSerSerAspSerPheIleThrSerThrGluTh 673
 4530 ACTCAGACA.....AATAGTCTGACTCAGACACATGTAACAAATCAGC 4573
 673 rThSerYrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeuMetL 690
 4574 ATTTCACATCATAGATCTCAAAAGAGATTCATTCAAACAAATTTTCTC 4623
 690 ysiLeSerAlaGlnAlaCysTyTrAsnGluSerIleHISGluLeuLysThr 706
 4624 ATGTTCAAGCA.....TCATCTACATTTATGACTTAAAGACA 4661
 707 Lys 707
 4662 AAA 4664
 seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:us-08-670-707A-5
 seq_documentation_block:
 ; Sequence 5: Application US/08670707A
 ; Patent No. 5859204
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; CITY: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/670,707A
 ; FILING DATE: 26-JUN-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/US94/13200
 ; FILING DATE: 15-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/212,133
 ; FILING DATE: 11-MAR-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/864,004
 ; FILING DATE: 07-APR-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Greenlee, Lorraine L.
 ; REGISTRATION NUMBER: 27,894
 ; REFERENCE/DOCKET NUMBER: 75-95F
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 303/499-8080
 ; TELEFAX: 303/499-8089
 ; INFORMATION FOR SEQ. ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 7493 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: cDNA to mRNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: repeat_unit
 ; LOCATION: 1..407
 ; OTHER INFORMATION: /rpl_type="terminal"

```

294 AlaIle.....AlaIleGlnSerGlySerLeuSerI 305
|||||
3561 GCATATTAAAGTCATATGATGATCA.....GAAGTAAAGC 3601
305 eLeuAlaGlyGlnSerValLeuPheGlnAsnSerGlnThrS 322
|||
3602 TTTCATTCATGATGACACCTTTTAAAGCAAAATTCACA..... 3641
322 eAspGlnGlyLeuValAlaGlnAlaIleTyrLeuGlnLysAlaIle 338
|||||
3642 .....TATTGACACTAAACCATATG 3662
339 LeuSer.....SerLeuGlnAlaArgAsnGlyAspIleLeuPhe..... 351
|||||
3663 CTAATATAGACACTCAACAAATAAAGACATATTTCATAGAAAGA 3712
332 PheAspProIleValGlnGlnSerSerIleSerProLeuProS 368
|||||
3713 TCAGATCTCTATTCACCAAGAGAGAAATACATCATGCCATTTCACA 3762
368 eSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerPro 384
|||||
3763 AGATCTTCTTCTGTCA.....CAATCTTCAAAW 3791
385 LeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGluAr 401
|||||
3792 TGCTTAAAGACCAATGGAATATTCCTTG..... 3824
401 GLeuSerGlnGluGlyThrProAspAsnLeuThrSerGlnLeuGlnG 418
|||||
3825 AACCTGACGAGAGACATAGTCCAAACCAATGATATTTTAAAG.... 3869
418 InProIleGlnLeuLysSerGlyArgLeuValLeuLysAspArgAlaVal 434
3870 .....TTTAAAAATATGTAAAAATCAAAAGCTTTC 3899
435 LeuSerAlaPro.....SerLeuSerGlnAspProGlnAlaIleLeu 448
|||||
3900 TTGTCAAGAAAAATTAAGTCCAGTAGACAGATGATTACAAAGAA 3949
448 uIleMetGlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysIleuA 465
|||||
3950 CATAGACTTAAAGACATGCCCTTTCCACATTAATATGACCAATTTTCCTAA 3999
465 IaThrLeuSerIleProLeuIleSer.....LeuAspThrLys 478
|||||
4000 CCACCTTGTCT...AACGTACATGAATATGAGCACAATCAAGAAAAA 4046
479 SerVal.....ThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
|||||
4047 AATATTTCAGAAAGATAGAGAAAGACACATAATTGAAGAAAGTAGT 4096
493 eLeuSerAsn.....SerGlyAspGlnAsnPheTyrGlnLysAsnV 506
|||||
4097 TTTCGCCCGAGTCACAGAACCACTGGCTTAAGAAATTTCTTGAAGAGCA 4146
506 aIGlnLeuLeuSerLysGlnGlnAsn..... 514
4147 TATTATACTAGGCACTAGCGCAAAATATATAGCTTTATATGACTACATGTA 4196
515 .....AsnIleProLeuLeuThrLeuSerLysGlnGlnSerHI 527
|||||
4197 CCAGTACTTCAAAACATCATCAATTAACANTCAACAAATACAGTAGCA 4246
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
|||||
4247 GATTACACATG.....GAGCATTTCTTTAAACAAGC. 4277
544 sPTTrpThrPheSerTrpLysAspSerAspGlnGlyHisSerLeuIleAla 560
|||||
4278 .....AAGGACAAGAAACAATTCAGACAGCTTGGTA 4310

```

```

561 AsnTrpThr.....ProLysAsnTyrValProHisProGlnAlaArgI 574
|||||
4311 AATAAACCAGAGAAATGTAAAAACATAT...CCAAAGCAGAGAAATAT 4357
574 nSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGI 590
|||||
4358 TACTACTACACGTAGTAAACGGCTTTGG.....GACAAATTC 4395
590 nAlaValGlnSerMetIleAsnThrIleAlaHisGlyValTyrLeu 606
|||||
4396 GACTGTCAACGTCAATGCGTTAAACCATTAACGTGTCAACACAGTGTATC 4445
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
4446 ATT..... 4448
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
|||||
4449 .....AACACATAGAC.....CACAGCAGCAAAATGAAAAGCTTC. 4484
640 nGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIa 656
|||||
4485 .....ATTACTAAATCTTCTCTATCAGATTCTCTGTGATTTAAAGCACC 4529
657 GlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
|||||
4530 ACTCAGACA.....AATAGTCTGACTCACACATTTGTAACATCAGC 4573
673 rThrSerTyrIleAlaThrValGlnAlaGlnIleuAlaThrSerLeuMetI 690
|||||
4574 ATTTCACCAATAGATGTCAAAAGAGCGCATTCCAAAACAATTTTCTC 4623
690 yIleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGlnLeuLysThr 706
4624 ATGTTCAGCA.....TCATCTACATTATGACTTTAAGACA 4661
707 Lys 707
4662 AAA 4664

```

seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-474-503-5

```

seq_documentation_block:
: Sequence No. 5744446
: Patent No. 5744446
: GENERAL INFORMATION:
: APPLICANT: Emory University
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: US
: ZIP: 30309
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,503
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Pratt, John S.
: REGISTRATION NUMBER: 29,476
: REFERENCE/DOCKET NUMBER: EMU106CIP(3)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6500
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 5:

```

```
284 pheleasnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerG1 300
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1886 ACTGGCGGCTCATATCATATATACCGGAATCTTACTGTTCAAAAG 1935
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
300 yGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1936 CCGTAACCTTCAAGCTATACAAATTTACACTTTTATGTAGCCGCTCA 1984
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
312 allPheGlnAsnAsnSer.....::: 318
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1985 .....TTTGACACACAAATGCGCTTCAACATTCCATTGCCAGAGAGG 2029
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
319 .....GlnArgThrSerAspGlnGlyLeuValArgAs 329
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2030 GCTAATTTAAAGATATCATATACACAGTAGCTTAATATATCCACCA 2079
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
339 nAlaIleTyrLeuGlyLysAspAlaIleLeuSerSerLeuGluAlaArg 346
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2080 CTGTGATACCACTTACCCACCATTTATAAGCAATATATCCACCAAAAT 2129
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
346 nGlyAspIleLeuPhePheAsp.....ProIle 355
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2130 CAGGTGATTTGAAATATATGATATAAAAGCGACGCTGAATCCAAAT 2179
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
356 ValGlnGlySerSerSerLysGlySerProLeuPro...SerSerLeuG1 371
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2180 GCGGCATATCTCTCAAAAGAGCAATCTCACAAATTTCTTCTGATTA 2229
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
371 nAlaSerValThrSerProThrProIleThrAlaSerProLeuValIleG 388
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2230 AGTAATATTTACCAAT.....C 2246
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
388 lnthSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGlu 404
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2247 AGATACATCAACAGCAGCGTTGCAAGGGGGCTTGTGATTCAGATGCG 2296
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
405 GluGluLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleG1 421
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2297 GCAGAAATGCT.....ACCTACTATTCAAACCAAA.....GA 2331
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
421 uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaValL 435
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2332 GTTAAATTTGGCAGCAGCACTTAATATTTTCAGCGCTTTAATAACCA 2381
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
435 euserAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleuGlu 451
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2382 TTACAGCTAAATAATGCGAGTGAT.....TTAACT 2410
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
452 AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr.... 466
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2411 ATTGGCATATGCTAGCGGTGATATGCTGCTATAAAAGTCACTTTGA 2460
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
467 .....LeuSerIleProLeuHisSerLeuAspThrG 477
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2461 CAAGCTTAACATTCAAAAATCTGCACGTACCGCTCACTCAATGTAAC 2510
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
477 lulySerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2511 ATAGCGAAGTG.....AAACGCTTAATGCTAGT..... 2539
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
494 leuSerAsnSerGlyAspLysAsnPheTyrGluAsnValGluLeuLeuSe 510
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2540 ...ACCATGCTGCTAATGATATAC...AGCACCGGTTTAAACCATTTCCG 2583
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
510 rlyGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2584 AAAAATATTAACGCTTAACAAATTAACGCT.....ACCTCCCA 2621
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
524 luginSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGly 540
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2622 AGACATAATATCTCTGCGCAGCAGAAATGTACAACCAAAAGAGC 2671
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
541 TyrGln...GlyAspTyrThrPheSerTyrLysAspSerAspGluGlyH1 556
```

```
2672 ACAATATCAATGCGCAACACAGCCGCTGGAAATACCTGCCAAATAG 2721
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
556 sSerLeuIleAlaAsnTyrThrProLysAsnTyrValProHisProGlu 573
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2722 TACAATTAAGCAACATTAACCTCCGCAAAATGTAACTGACAGCAACAG 2771
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
573 rGlnSerThrLeuValAlaAsnThrLeuTyrAsnThrTyrSerAspMet 589
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2772 AAATCTGTGTACACAGAGATGCTGTATTAATGACCAACCGCGGCA 2821
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLe 606
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2822 GTAACATTAATGCAAAACAGCGGATGTAAAGTGCA.....AT 2862
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
606 uPheGlyThrTrpGly.....SerAlaValSerAsnLeuPheT 619
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2863 TCAATCAACTCCCGTAATGTAAATATTAACAGCGCAGCAATACACTTA 2912
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
619 yAlaHisAspSerSerGlyLysProIleAspAsnThrHisArgSer 635
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2913 AGGTAGTAATATCACTGTGCAAGTGTA..... 2941
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsnHisSerPh 652
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2942 .....ACAGTAACAGCGCATGCGACGACCTT 2967
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
652 eCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspPheIleT 669
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2968 GACACGTACAGCAGCGCTCAACCATTAAGTCGACACAGCAATGCCAAATA 3017
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeu... 684
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3018 TTACACCAAAACAGGTATATCAACGGTAAGTTGAATTCACCTCCGCG 3067
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
685 AlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlnSerI1 701
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3068 TCTGTAAACACTTGTTCAGCACTGACCACTCTGCTGTA..... 3106
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
701 eHisGluLeuLysThrLysTyrArgSerPheSerLysGluGlyPheGlyS 718
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3107 .....GTAATATTT 3116
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
718 eTrpHisSerValAlaVal.....SerGlyGluValCysAlaSer 731
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3117 CAGGTAAACACTGTTACTATCTACGCGGATAGCGGAATTAACCTCCCA 3166
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
732 lIleProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePh 748
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3167 GTAGGTCTTACAATTAATGCGACTAATAGTGAACCACTCAAGCCAAATC 3216
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
748 eSerLysLeuGlnGly...PheSerGlyThrGlnAspGlyPheGluGluS 764
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3217 AGCGCATATGTAAAGGTACAAATTTCTGTGAATTAACAGTAATGTTCAC 3266
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
764 eSerGlyGluIle 768
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3267 GCACGTGCTGATTTA 3280
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
seq_name: /cgn2.6/prodata/2/ina/5a_COMB.seq:us-08-212-133A-7
seq_documentation_block:
; Sequence 7, Application US/08212133A
; Patent No. 5663060
; GENERAL INFORMATION:
; APPLICANT: Lohr, John S.
; APPLICANT: Runge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
```

170 SerSerLysGlyAsnGlyValAlaIleGlnAlaGlnThrPheSerLeuSe 186
1448 ATTGAGTCTGAGGGGGGGTACGCTTAATTCAAACTTAACGCTCATC 1497
186 fArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
1498 TAGCAACATACAAACCCCTGGCGTAATTATAAATCTCAAAACCTTAATG 1547
201 euAsnGlyValAlaIleGlyCysSerAsnLeuIleCysSerGlyAsnVal 217
1548 TCTCAGAGGGTCAACTTA.....AATCTCAAGGCTGAAGGTTCAACA 1591
218 AsnProLeuPhe.....PheThrGlyAs 225
1592 GAACCGCTTTTTCATAGAAAATGATTTAAACTTAACGCCACCGGTGG 1641
225 nSerLalThr..... 228
1642 CAATTAACAATCAGACAAGTCGAGGTACCGATTACCGCTCAACAAG 1691
229AsnGlyValAlaIleGlyCys 235
1692 GTGTGGCAGCCAAAAAACATACCTTTAAAGGGGTAATATTCACCTTC 1741
236 IleSerAspLeuAsnThrSerGlu...LysGlySerLeuSerLeuAlaCyl 251
1742 GGCTCTCAAAAAGCCACACAAGAAATCAAGGCATGTTACCATCAATAA 1791
251 sAsn...GlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGlyG 267
1792 AAMACTAACCTACTCTTCGTGTGCGAATTTTCCGCAAAACAAATCGC 1841
267 LyAlaIleThrAlaLysHisMetValLeuArgTyfAsnGlyProValSer 283
1842 CTTTAATATATACAGGAATGTTATT.....AATAATGGCACCTTACC 1885
284 PheIleAsnAsnSerAlaLysIleGlyValAlaIleGlnIleGlnSerG 300
1886 ACTGCCGCTCATATCAATATACCGCAAAATCTTACTGTTCAAAAG 1935
300 yGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
1936 CGCTAACCTTCAAGCTATACAAATTAACCTTTAATGTAGCCGCTCA. 1984
312 AlLeuPheGlnAsnAsnSer..... 318
1985TTTGACACAAATGGCGCTCAACATTTCCATGCCAGAGAGAG 2029
319GlnArgThrSerAspGlnGlyLeuValArgAs 329
2030 GCTAATTTAAAGATATCAATATACACAGTACCTTAATATATACACCA 2079
329 nAlaIleThrLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArgA 346
2080 CTCGTATACCACTTACCGCACCATTTATAAAGCAATATATCCAACAAT 2129
346 sncIAspIleLeuPhePheAsp.....ProIle 355
2130 CAGGCGATTGATATATATGATAAAAAAGCGCATGAATTCGAAT 2179
356 ValGlnGluSerSerSerLysGluSerProLeuPro...SerSerLeuG 371
2180 GCGCGCAATATCTCAAAAAGCAAGCAATCTCACAAATTTCTTCATGA 2229
371 nAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIleG 388
2230 AGTAATATTTACCAAT.....C 2246
388 InThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGlu 404
2247 AGATAAACATCAAAAGCGCTTGAAAGGGGGCTTCTGATTCAGAGTGAG 2296

405 GluGluLysThrProAspAsnLeuThrSerGlnLeuGlnProIleG 421
2297 GCAGAAATGCT.....AACCTAATCTATTCAAACCAAA.....GA 2331
421 uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaValL 435
2332 GTTAAATTTGGCAGAGACCTAAATATTTTCAGGCTTTAATTAACGAGAA 2381
435 euSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGlu 451
2382 TTACAGCTAAATAATGGCAGTAT.....TTAACT 2410
452 AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr.... 466
2411 ATTGGCAATCTAGCGGTGTGATATGCTGATCTAATAAAGAGACTTTGA 2460
467LeuSerIleProLeuHisSerLeuAspThrG 477
2461 CAAGGTTAAGATTCCAAAATCTCGACTGAGCGGTCCACATATGACACTA 2510
477 LysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
2511 ATACCGAAGTC.....AAACGTTAATGCTACT..... 2539
494 LeuSerAsnSerGlyAspGluAsnPheTyfGluAsnValGluLeuLeuSe 510
2540 ...ACCAATGCGGTGAATGATAAC...AGCACCGGTTAACCATTTCCGC 2583
510 fLysGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
2584 AAMAGATTAACGGTAAACAAATACGTT.....ACCTCCCA 2621
524 LucGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheG 540
2622 AGACATTAATATCTCTGCCAGCAGCAAAATGTACACCAAGCAAGAGC 2671
541 TyfGln...GlyAspThrPheSerThrPheLysAspSerAspGluGlyH 556
2672 ACAACTATCAATGCCAACACAGCAGCGCTGGAAGTAAGTGCCTCAAAATG 2721
556 sSerLeuIleAlaAsnThrPheProLysAsnTyfValProHisProGluA 573
2722 TACAATTAAGCAACATTAACCTCGCAAAATGTAACTGACAGCAACAG 2771
573 rgGlnSerThrLeuValAlaAsnThrLeuThrPasnThrTyfSerAspMet 589
2772 AAAATCTGTACCCACAGACAAATGCTCATTAATGCAACACCGCGCA 2821
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyfLe 606
2822 GTAAACATTAGACAAAAAGCGGATATTAAGGTGGA.....AT 2862
606 uPheGlyThrTrpGly.....SerAlaValSerAsnLeuPheT 619
2863 TGAACTCAACTTCGGTAATGTAATATTTACAGCGGCAATACACTTA 2912
619 yAlaHisAspSerSerGlyLysProIleAspAsnThrHisArgSer 635
2913 AGGTAGATTAATATCTGCTCAAGATGA..... 2941
636 LeuGlyTyfLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
2942ACAGTAACACCGGATGACAGAGCCTT 2967
652 ecysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspPheIleT 669
2968 GACAACCTACAGCAGGCTCAACCATTAGTGCGCAACAGCAGCAATAA 3017
669 hrSerThrGluThrSerTyfIleAlaThrValGlnAlaGlnLeu... 684
3018 TTACAACCAAAACAGCTGATATCAACGGTAAGTTGAATCCAGCTCGGC 3067
685 AlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyfAsnGluSerI 701

MOLECULE TYPE: DNA (genomic)
US-08-530-198-7

alignment_scores:

Quality: 144.00 Length: 788
Ratio: 0.364 Gaps: 38
Percent Similarity: 50.254 Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-530-198-7

Align seg 1/1 to: US-08-530-198-7 from: 1 to: 4287

```
72  GlysSerThrCysTrpPheValSerLeuHisIleThrAspProly 88
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1208  GGAGCCAGCTAC...TGCAGCTAACCACTTAAATGTACCTGC... 1249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
88  sgluAlaLeuPheLysGluLysGlyAspLeuSerIleGlnAsnPheArg 105
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1250  .....GCTAGTAAATTTAACTCTCCATTCAGCAGC... 1279
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
105  heLeuSerPheThrAspCysSerSerLysGlyUserSerProSerIleIle 121
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1280  .....ACAGGAAGTGGCTCAACAGGTCAGCAGCTA... 1309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
122  HisGlnLysAsnGlyGlnLeuSerLeuArgAsnGlySerMetSerPh 138
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1310  ....CCCATGCGAGATTAATGCCATTAATTAATAAGCCACTTT 1353
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
138  eCysArgAsnHisAlaGluGlySerGly...AlaIleSerAla 153
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1334  T.....AATATCGCACAGCCCTCAACAGCTAATCTTACATCAAGCCAT 1397
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
153  sPalaPheSerLeuGlnHisAsnTyrIlePheThrAlaPheGlnLysn 169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1398  CAATATATCCCTTTAAGAGTAACGCTAAGCTATTAATTAATGAGAT 1447
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
170  SerSerLysGlyAsnGlyGlyAlaIleGlnIleGlnThrPheSerLeu 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1448  ATTTGAGTCTCAGCGCGGCGGAGCTTAATTTCAACTTAACGCCCTCATC 1497
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
186  rArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAsp 201
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1498  TACCAACATACAAACCCCTGCGCTAATTAATAAATCTCAAAACCTTAATG 1547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
201  euAsnGlyGlyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1548  TCTCAGGAGGCTCACTTA.....AATCTCAAGGCTGAAGCTTCACCA 1591
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
218  AsnProLeuPhe.....PheThrGlyAs 225
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1592  GAAACCCGCTTTTCAATAGAAATGATTTAAACCTTAACGCCACCGCTGCG 1641
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
225  nSerAlaThr..... 228
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1642  CAATATACCAATCAGACAGTGCAGCGATTCACGCCGTCAACAAG 1691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
229  .....AsnGlyGlyAlaIleCysCys 235
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1692  GTGTGCGACCCAAACAAACATTAATTTAAAGGGGTAATATACCTTC 1741
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
236  IleSerAspLeuAsnThrSerGlu...LysGlySerLeuSerLeuAlaCys 251
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1742  CGCTCTCAAAACCAACCAAGAAATCAAGAGCATGTATACCATCAATTA 1791
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
251  sAsn...GlnGluThrLeuPheAlaSerAsnSerAlaLysGlyGlyG 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1792  AAACACTAAGCGTACTCTTCGTGCGAATTTTCCGAAACCAATATCC 1841
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
267  LyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSer 283
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1842  CTTTAATATATAGCAGAAATGTATT.....AATATAGCACTTAC 1885
```

```
284  PheIleAsnAsnSerAlaLysIleGlyGlyAlaIleAlaIleGlnSerG 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1886  ACTGCGCGCTCATTAATATATACAGCGGAATCTTACTGTTCACAAAG 1935
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
300  yGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1936  CGTAACCTTCACGCTAATACAAATTAACCTTTAATGTACCGGCTCA 1984
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
312  aLeuPheGlnAsnAsnSer..... 318
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
1985  ....TTTGACACAAATGCGCTTCAACATTTCCATTGCCAGAGAGG 2029
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
319  .....GlnArgThrSerAspGlnGlyLeuValArgAs 329
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2030  CCTAATTTAAGATATATCAATACACCGTAGCTTAATATATACACCA 2079
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
329  nAlaIleTyrLeuGlnLysAspAlaIleLeuSerSerLeuGlnAlaArg 346
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2080  CTCTGATACCACTTACCGCACCATTTAAGCAATATATCAACAAAT 2129
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
346  snglyAspIleLeuPheAsp.....ProIle 355
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2130  CAGGTGATTTGAATATTTATGATTAATAAAGCAGCGTGAATCCAAAT 2179
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
356  ValGlnGluSerSerSerLysGlySerProLeuPro...SerSerLeuG 371
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2180  GCGCGCATATCTCTCCAAAGAAAGGCAATCTCAATTTCTTGATGA 2229
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
371  nAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIleG 388
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2230  AGTAATATTTACCAAT.....C 2246
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
388  lInThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2247  AGATATACATATCAAGCAGCGGTGAAGGGGCGCTTGATTAAGTAG 2296
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
405  GluGluLysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleG 421
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2297  GCAGAAATGCT.....AACCTAATCTTTCACAAACCA...GA 2331
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
421  uLeuLys...SerGlyArgLeuValLeuLys.....AspArgAlaVal 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2332  GTTAAATATGACAGAGACCTAATATTTCAAGCTTTAATAAGCAGAA 2381
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
435  euSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetG 451
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2382  TTACAGCTAAATAATGCGAGTCT.....TTAAGCT 2410
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
452  AlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr.... 466
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2411  ATGGCAATGCTAGCGGTGTAATGCTGATGCTAATAAAGTACCTTTGA 2460
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
467  .....LeuSerIleProLeuHisSerLeuAspThrG 477
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2461  CAAGGTAAAGATTCAAAATCTCGACTGACGGTCAACATGTACACTTA 2510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
477  LuLysSerValThrIleHisIleProAsnLeuSerIleGlnLysIlePhe 493
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2511  ATAGCAAGTGC.....AAACGCTAATGCTAGT..... 2539
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
494  LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeu 510
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2540  ...ACCATGCTGTAATGATAAC...AGCACCGGTTTAACCATTTCCG 2583
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
510  rLysGlu.....GlnAsnAsnIleProLeuLeuThrLeuSerLysG 524
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2584  AAAAGATATTAACGGTAACAAATACGTT.....ACCTCCCA 2621
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
524  LuGlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheG 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
2622  AGACAATAATATATCTCGCGCAGCAGGAATATGAACAACCAAGAACG 2671
```

seq documentation block:

```

; Sequence 7, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-7

```

alignment scores:

```

Quality: 144.00 Length: 788
Ratio: 0.364 Gaps: 38
Percent Similarity: 50.254 Percent Identity: 19.670

```

alignment block:

```
US-09-677-752-4 x US-08-302-832-7
```

```
Align seg 1/1 to: US-08-302-832-7 from: 1 to: 4287
```

```

72 GlnAsnSerTyrCysTirPheValSerLysLeuHisIleThrAspProLys 88
|||||
1208 GGAGCAGCCTAC...TGGACGTAACCACTTAAATGTTACTG... 1249
88 sGluAlaLeuPheLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||
1250 .....GGTAGTAATTAACCTCTCCATTGACAGC..... 1279
105 heluSerPheThrAspCysSerLysGlySerSerProSerIleLle 121
|||||
1280 .....ACAGGAAGTGCCTCAGCAGTCCAGCAAGTA... 1309
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||||

```

```

1310 .....CGCAATGCGAATTAAATGGCATACATTAAATGAAGCCACTT 1353
138 eCysArgAsnHisAlaGluGlySerGlyGly.....AlaIleSerAla 153
|||||
1354 T.....AAATGCGACAGAGGCTCAGACGTAACCTTAAATGAGCAT 1397
153 spAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheGluGlnAsn 169
|||||
1398 CAATATATCCCTTTAAGAGTAACGCTAAGTACGATATTAATGAAGAT 1447
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSerLeu 186
|||||
1448 ATTCAGTCTCAGGGGGGGGCTTAAATTCACACTTAACGCTCATC 1497
186 ArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaSpl 201
|||||
1498 TAGCAACATCAAAACCCCTGGCGTAATTAATAATTCGAAACCTTAATG 1547
201 euAsnGlyGlyAlaIleCysSerAsnLeuIleCysSerGlyAsnVal 217
|||||
1548 TCTCAGGAGGCTCAACTTA.....AATCTCAAGGCTGAAGGTTCAACA 1591
218 AsnProLeuPhe.....PheThrGlyAs 225
|||||
1592 GAAACCGCTTTTCAATAGAAATGATTAACTTAACGCCACCGGTGG 1641
225 nSerAlaThr..... 228
1642 CAATATACAAATCAGACAGTGCAGGGTACCGATTACGCTCAACAAG 1691
229 .....AsnGlyGlyAlaIleCysCys 235
|||||
1692 GTGTCGCGACGCAAAAACATTAACCTTTAAAGGCGGTAATATACCTTC 1741
236 IleSerAspLeuAsnThrSerGlu.....LysGlySerLeuSerLeuAla 251
|||||
1742 GGCTCTCAAAAAGCCACACAGAAATCAAAAGCAATGTTACATCAATAA 1791
251 sAsn.....GlnGluThrLeuPheAlaSerAsnSerAlaLysLysGlyG 267
|||||
1792 AAACACTACCTCTCTCTCTGCGGCAATTTGCGCAAAACAAATTCGC 1841
267 LysAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSer 283
|||||
1842 CTTTAATATATAGCAGCAATGTATTT.....AATATGCAACCTTACC 1885
284 PheIleAsnAsnSerAlaLysIleGlyGlyAlaIleAlaIleGlnSerG 300
|||||
1886 ACTGCGGCTCCATTATCAATATAGCCGGAATCTTACTGTTCAAAAGG 1935
300 YGlySerLeuSerIleLeuAlaGly.....GluGlySerV 312
|||||
1936 CGCTAACCTTCAACCTATATACAAATTTACACTTTTAATGTACCGGCTCA 1984
312 AlLeuPheGlnAsnAsnSer..... 318
|||||
1985 .....TTTGACACACATGCGCGCTTCAACATTTCATTCAGCAGAGAGG 2029
319 .....GlnArgThrSerAspGlnGlyLeuValArgAs 329
2030 GCTAATTTAAAGATATCAATATACACCGTAGCTTAATATATACACCAA 2079
329 nAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLeuGluAlaArg 346
|||||
2080 CTCTGATACCACTTACCGCACCATTATATAAAGCAATATATACCAACAAT 2129
346 snGlyAspIleLeuPhePheAsp.....ProIle 355
|||||
2130 CAGCTGATTTGAATATATGATTAATAAAGAGCAGCGTGAATCAAAATT 2179
356 ValGlnGluSerSerLysGlySerProLeuPro...SerSerLeuG 371
|||||
2180 GCGGCAATATCTCACAAAAGAGCAATCTCACAATTTCTTGATATA 2229

```

```
662 LysSerSerAspSerPheIleThrSerThrclutrhThrSerTyrIleAl 678
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2023 ACCTGGGTAGTAATTTAACTCTTCATTCAGACACAGGAGGGCTC 2072
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
678 aThValGlnAlaGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnA 695
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2073 AACAA..... 2076
695 lacYtYrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPhe 711
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2077 .....GGTCCAGACATACCGAATGAGAAATTAATGGCTAATACATT 2118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
712 SerLysGluGlyPheGlySerThrPheSerValAlaValSerGlyGluVa 728
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2119 AATAAGCACCTTTAATATATCGCACAGGCTACAGCACTAATTAGCAT 2168
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
728 lCysAlaSerIle...ProIleValSerAsnGlySerGlyLeuPheSerS 744
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2169 CAAAGCATCAATTAATGCTTTAAAGCTAAC.....GCTA 2203
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 ePheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGly 760
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2204 ACTAGCATTAATTATATGAGATATTTCAGTCTCAGGG..... 2241
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 pheGluGluSerSerGlyGluLeuLeuArgSerPheSerAlaSerSerPhe 777
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2242 .....GGGGGTAGCCTTAATTTCAACTAAGCCTCACTCTAG 2279
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 gAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLysT 794
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2280 CAACATACAAACCCCT...GGCGTAATTATA.....AATCTCAAAAC 2319
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
794 hArgThrTyrTyrPheLeuGlyAlaTyrIleGlnAspLeuLysArg 810
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2320 .....TTTAATGCTCTCAGAGGGGTCACTTAATCTCAAGGCT 2358
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
811 AspValGluSerGlyProValIleLeuLeuLysAsnAlaValSerTyrAs 827
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2359 GAAGGTTCACAGAAACCCCTTTTCATATAGAAATGATTTAACTTA 2408
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 pAlaProMetAlaAsnLeuAspSerArg 836
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2409 CCCACCCGCTGCCAATATTAACAATGACA 2436
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/prodata/2/lna/5A_COMB.seq:US-08-038-682-7
seq_documentation_block:
: Sequence 7, Application US/08038682
: Patent NO. 5549897
: GENERAL INFORMATION:
: APPLICANT: BARENKAMP, STEPHEN J
: APPLICANT: ST. GENE III, JOSEPH W
: TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Maltare, Ltd
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/038.682
: FILING DATE: 16-MAR-1993
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
```

```
NAME: BERKSTRESSER, JERRY W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-293
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-038-682-7

alignment_scores:
Quality: 144.00 Length: 788
Ratio: 0.364 Gaps: 38
Percent Similarity: 50.254 Percent Identity: 19.670

alignment_block:
US-09-677-752-4 x US-08-038-682-7

Align seg 1/1 to: US-08-038-682-7 from: 1 to: 4287

72 GlysAnsSerTyrCysTrpPheValSerLysLeuHisIleThrAspProly 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1208 GGAGCGACCTAC...TGAAGCTAACCACTTAATGTGTACTCTG.... 1249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 sGluAlaLeuPheLysGlyLysGlyAspLeuSerIleGlnAsnPheArgp 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1250 .....GGTAGTAATTTAACTCTCCATTGACACG..... 1279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 heLeuSerPheThrAspCysSerLysGluSerSerProSerIleIle 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 .....ACAGCAAGTGCCTCAACAGCTCCAAAGCATA... 1309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 HisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPh 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1310 .....CGCAATGCAGAAATTAAATGCGCATACATTATTAAGCCACTTT 1353
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 eCysArgAsnHisAlaGluGlySerGlyGly....AlaIleSerAla 153
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1354 T.....AATATCGCAACAAAGCTCAACAGCTACTTAGCATCAAGGAT 1397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 CAATAATGCCCTTTAAGAGTAGCTTAATTTCAACTACGCAATTATTAAGA 1447
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 SerSerLysGlyAsnGlyGlyAlaIleGlnAlaGlnThrPheSerLeu 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1448 ATTTAGTCTCAGGGGGGTAGCGTTAATTTCAACTTAACGCTCATC 1497
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 fArgAsnValSerPro.....IleSerPheAlaArgAsnArgAlaAspL 201
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 TAGCAACATACAAACCCCTGGCGTAAATTAATAATCTCAAACTTAATG 1547
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 euAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVal 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1548 TCTCAGAGAGGTCAACTTTA.....AATCTCAAGCTGAAGAGTTCAACA 1591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218 AsnProLeuPhe.....PheThrGlyAs 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 GAACCCGCTTTTCATATAGAAATGATTTAACTTAAGCCACCGGTGG 1641
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
225 nSerAlaThr..... 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1642 CAATATACAAATCAGACAAAGTGAGGTAACGATTCACGGCTCAACAAG 1691
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
229 .....AsnGlyGlyAlaIleCysCys 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1692 GTCTCGACACCAAAAAAACAATAACTTTAAAGCGGTAAATATCTACCTTC 1741
```



```

387 11eclInhrSerAlaAsnArgSerVal11ephSerSerGluArgLeuSe 403
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2043 GTGAGACATCGGGCATTTATTTATTCATTCGACGACATGCAATGTTAA 2092
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
403 rglGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGng 418
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2093 AACAAAGAGGGTGGCTAGACCCCTGATGATGTACCAATTTGACGCCGAG 2142
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
418 InProLeuGluLeuLysSerGluArgLeuValLeuLysAspArg..... 432
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2143 ACCCTTCGCAATATATACCGGT.....ATAAATGATGAAATTCGCCA 2183
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2184 ACAGGACCCGGTGAAGCAACGACCTTAAAAAATAGCAATCAAAAC 2233
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
446 AleuLeuLe.....:|||||:|||||:|||||:|||||:|||||:|||||
2234 AACGCTAACCAATACAACTATTCTGAATTTCTGAANAACCGCTGGACAA 2283
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
450 .....MetGluAlaGluThrSerLeuLysThrSerAspLeuLysLeu 464
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2284 TGAATATACGCGCATCAAGAAACTTACCGCTTAATAGCTCATCAATC 2333
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
465 AlaThrLeuSer...IleProLeuHisSer..... 473
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2334 CGAAGCAACTCCACCTTAATCTCCATAGTAAGGTCAGCGTGGCCGACG 2383
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 .....:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2384 CGTTACATTTATGAGATATTACTTCTAAGCGGCAATTTAACCATTT 2433
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2434 ATTCGGCGGATGGGTGATGTTCTATAAATATACCTTGATCAGGGT 2483
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGluAspGluAspH 502
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2484 TTTTAAATATTACCGCGCGCTTCGCTAGCTTTTGAAGGTGAATAACAA 2533
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
502 eTyGlu.....AsnValGluLeuLeuSerLys..... 511
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2534 AGCAGCGCAGCGGCAATGCTAAATGTGCGCCAGGCGACCTGAACCA 2583
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
512 .....GluGlnAsnAsnIleProLeu..... 518
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2584 TTACAGAGAGGAAAGATTTCAGGGCTAACAGCTATCTTTAAACGGA 2633
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
519 .....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2634 ACGGCTAAAGTCTGAATATCATTTTCATTCAGTGAATATTTTA..... 2675
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
531 GAspGluAsnLeuSerSerHisPheGluTyrglnLysAspThrPheS 548
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2676 .ACCCCAATCTTAGTGGCACAATTAACATATCTGGAAATATACAAATTA 2724
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
548 eT.....TrpLysAspSerAspGluGly 555
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2725 ACCAACTACGAGAAAGAACACCTGATTTGGCAACACCATGATTCG 2774
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
556 HisSerLeuLe.....AlaAsnTrpThrPr 564
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2775 CACTGCAAGCTCAGTCTCTTAATCTAGACACAGCGCCAAATTTTACCTT 2824
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
564 OlysAsnTyValProHisProGluArgLysSerThr..... 576
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2825 TATTAAATACATTTCAAGCAATAGCAAGGCTTACACACAGTATAGAA 2874
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
576 ..... 576
2875 GCTCTGACAGGGTGAATTTTAAAGCGGTAATGTCACATGTCATTCAT 2924

```

```

577 .....LeuValAlaAsnThrLeuTr 583
2925 CTCAAACAGACAGCAAGATTAAATTTCAAATTTAAACCAACGAGACAT 2974
583 PAsnThrTySerAspMetGlnAlaValGlnSerMetIleAsnThrIleA 600
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2975 GAACACA...AGCAACCTTACCAATATCGGTTTATGCCAATATCAAG 3021
600 IAnSglGlyAlaTyLeuPheGluThrTrpGlySerAlaValSerAsn 616
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3022 CCACGTGCTGGC.....GGCTCTGTTTTTTTCAT 3050
617 LeuPheTyAlaHisAspSerSerGlyLysProIleAsp...AsnTrpH 632
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3051 ATATATGCCAACCATTTCTGCGACAGGGCTGACGTTAAATGACGTGAAT 3100
632 SHisArgSerLeuGluTyLeuPheGluIleSerThrHisSerLeuAspA 649
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3101 TAATATCTTACAGCGCGCTAATTTTACTTAATTTCCATGTTGCGGCG 3150
649 sPHisSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAsp 665
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3151 ATGACGCTTTTAAATC.....AACAAAGAC 3176
666 SerPheIleThrSerThrGluThrThrSerTyrlleAlaThrValGlnAl 682
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3177 TTAACCAATAATGCCAACCAATTC..... 3200
682 aGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyraSng 699
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3201 .....AAT 3204
699 IuserIleHisGluLeuLysThrLys...TyraArgSerPheSerLysGlu 714
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3205 TCAGCGCTCAGACAGACAGAAAGATGATTTTATGACGGGTACGACGCAAT 3254
715 GlyPheGlySerTrpHisSerValAlaValSerGly 726
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3255 GCCATCAATTCACCTTACACATATTCATCTCGGCG 3290
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-617-697-6
seq_documentation_block:
; Sequence 6, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5977336-Typhable Haemophilus
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/617,697
; FILING DATE: 01-Apr-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 05-Oct-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993

```

```

1656 GATTCTGAAGCAAGATAAAGCGCATATTGTTCTTCCGCCAAGA 1705
273 smevValLeuArgTyrAsnGlyProValSerPheIleAsnSerIleAl 290
1706 GCGTGAAGCGGAAATTGGCGGTAAATTTCGCTCAAAATCAGCAACTA 1755
290 ysiIleGlyValIleAlaIleGlnSer..... 299
1756 AA...GGCGGCAAGCTGATGTAAGTCCGATAAGTCACATTAAACA 1802
300 GlycylSerSerIleLeuAlaGlyGlySerValLeuPheGlnAs 316
1803 GGTGCACATTATCGACCTTTGAGTAAAGGGGAGCAAACTTACCTTGC 1852
316 naenSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
1853 CGGTGACGAGCGC.....GGCGAAGTAAACGGCGCATTCATC 1890
333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
1891 TAGCAAGAAACCTCTTAGAAAAGGCTCAACCATCATGTTACAGGC 1940
347 .....GlyAspIleLeuPheAs 353
1941 AAAGAAAAGCGGACCGCCATTTGTGTGGGCGCATATTGGCTTAATTGA 1990
353 proIleValGlnLysSerSerLysGlySerProLeuProSerSerL 370
1991 C.....GGCAATA 1998
370 euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
1999 TTACGCTCAAGTAGTGTGATATCGCTAAACCGGTGGT.....TTT 2042
387 IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
2043 CTGGAGACATGGGGCATTTTATTCATTCAGACCAATGCAATGTGTAA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
2093 AACAAAAGAGTGGTGTGACAGCCGTGATGATGATGAAATGAAAGCCGAG 2142
418 InProIleGluLeuLysSerGlyArgLeuValIleuLysAspArg..... 432
2143 ACCCCCTTCGCATATATACCGGT.....ATAATGATGATATTCCTCA 2183
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
2184 ACAGGCACCGGTGAGCAACGACCAACCTAAAAAAATACGAGACGCAAAAC 2233
446 aleuLeuIle..... 449
2234 AACGTAACCAATACACATATTCAATTAATCTGAAAAACGCCCTGAGCA 2283
450 ....MecGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
2284 TGAATATATACGGCATCAGAAAACTTACCCTTAATAGCTCATCAACATC 2333
465 AlaThrLeuSer...IleProLeuHisSer..... 473
2334 GGAAGCAACTCCCACTTATATCTCATAGTAAGTCAACCGTGGCGAGG 2383
473 ..... 473
2384 CCTTCAGATTGATGAGATATTACTTTAAAGCGGAAATTTAACCATTT 2433
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
2434 ATTTCGGCGGATGGGTGATGTTCAATAAATATTACGCTTATCAGAGGT 2483
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
2484 TTTTAAATATTAACCGCCCTTCGTAAGCTTTTGAAGGTGCAAAATACAA 2533

```

```

502 eTyrGlu.....AsnValGluLeuLeuSerLys..... 511
2534 AGCAGCGGAGCGGCAAAATCTTAATTTGTCGCCAGGCGACTGTACCA 2583
512 .....GluGlnAsnAsnIleProLeu..... 518
2584 TTACAGAGAGAGGAAAAATTTCAGCGCTAACAGCATCTTTAAACGCA 2633
519 .....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuPr 531
2634 ACGGTAAAGGTCTGAATATCATTCATTCATGAGTAATTA..... 2675
531 oAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPheS 548
2676 .ACCCACAAATCTTAGTGGCAATTAACATATCTGGAAATATACATTA 2724
548 er.....TyrPlysAspSerAspGluGly 555
2725 ACCAACTACAGAGAAAGACACCTCGTATTTGGCAACACCATGATTCG 2774
556 HisSerLeuIle.....AlaAsnThrPhePr 564
2775 CACTGGAACGTCACTGCTTAATCTAGAGACAGCCGCAATTTTACCTT 2824
564 oLysAsnTyrValProHisProGluArgGlnSerThr..... 576
2825 TATTAAATACATTTTCAAGCAATAGCAAAAGCCTTAACACACAGATAGAA 2874
576 ..... 576
2875 GCCTGACAGGGTGATTTTAAACGGCTAAATGGCAACATGTCAATCAAT 2924
577 .....LeuValAlaAsnThrLeuTr 583
2925 CTCAAAGAGAGCGGAAAGTTAATTTCAATTAATAACCAACGAGAACAT 2974
583 PAsnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleA 600
2975 GAAACAA...AGCAAACTTACCAATTTGGTTTACCAATATCAACG 3021
600 IAniGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsn 616
3022 CCACGTGGTGG.....GGCTCGTTTGTGAT 3050
617 LeuPheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnThrPhI 632
3051 ATATATGCCAACCATTTCTGGCAGAGGGCTGAGTTAAATGAGTGAAT 3100
632 SHisArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspA 649
3101 TAAATCTCTAACGGCGCTAATTTACTTAAATTCCTATGTCGGCGG 3150
649 sPHisSerPheCysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerAsp 665
3151 ATGACGCTTTTAAATC.....AACAAAGAC 3176
666 SerPheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAl 682
3177 TTAACCAATAATGCAAACCAATTCA..... 3200
682 aGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnG 699
3201 .....AAT 3204
699 LuSerIleHisGluLeuLysThrLys...TyrArgSerPheSerLysGlu 714
3205 TCAGCCTCAGACAGAGAAAGATGATTTTATGACGCGTACGACGCAAT 3254
715 GlyPheGlySerThrHisSerValAlaValSerGly 726
3255 GCCATCAATTCACCTACACATATTCATTCGGGC 3290

```

```
1415 :|||||:.....: |||||:
CGCAGTGTAAATCTATT......GTGCAAG 1443
150 LeSerAlaSpAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
:....:
1444 TGAA..... 1448
167 GlnGluAsnSerSerLeuGlyAsnGlyAlaAlaIleGln.....Al 180
:|||||:.....: |||||||:.....:
1449 AACGAGCGCTGATTAGCGTAAATGCTGCGACATTTCTTACTCCGACG 1498
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
:|||||:.....: |||||:.....:
1499 GCAAATAATACCATCACCGATATATTAACCAACCATTTACTTACAGCA 1548
192 LeSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaAlaIleCysGly 208
:|||||:.....: |||||||:.....: |||||:.....:
1549 TTGCGCGCCCTGAAAATGAGACG...GTCAATCTGGCGGATATTTTCC 1595
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
:|||||:.....: |||||||:.....: |||||:.....:
1596 AAA.....GCCGCTACATTAATGTC.....CG 1618
225 nSerAlaThrAsnGlyAlaAlaIleCysCysIleSerAspLeuAsnThrS 242
:|||||:.....: |||||:.....:
1619 TGCTGCCACT..... 1628
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrIleuPheAla 258
:|||||:.....: |||||:.....: |||||:.....:
1629 .....ATTGAAACCAAGTAACTTCTGCT 1655
259 SerAsnSerAlaLysGluLysGlyAlaAlaIleTyr.....AlaLysIle 273
:|||||:.....: |||||||:.....: |||||:.....:
1656 CATCTGTAGCAAGAAAGATAAAGCGCAATATTTCTTCCGCCAAACA 1705
273 smeValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
:|||||:.....: |||||||:.....: |||||||:.....:
1706 GCGTGAACGCGAAATTTGGCGGTATTTCCGCTCAAAATCAGCAAGCTA 1755
290 ysIleGlyAlaAlaIleAlaIleGlnSer..... 299
:|||||:.....: |||||||:.....: |||||:.....:
1756 AA...GGCGCAAGCTGATGATNAAGTCCGATAAAGTACATTAAACA 1802
300 GlyGlySerLeuSerIleLeuAlaGlyLysIleSerValLeuPheGlnAs 316
:|||||:.....: |||||||:.....: |||||:.....:
1803 GGTGAGATTATCGACCTTTCAAGTAAGAGGAGGAGAACTTACTCTTGG 1852
316 nasnSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrL 333
:|||||:.....: |||||:.....: |||||||:.....:
1853 CGGTGACGAGCGC.....GCCGAAGCTAAACCGCATTCAT 1890
333 euGluLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
:|||||:.....: |||||||:.....: |||||:.....:
1891 TAGCAAGAAACCTCTTAGAAAAAGCTCAACCATCATGTATTCAGGC 1940
347 .....GlyAspIleLeuPhePheAs 353
:|||||:.....: |||||||:.....: |||||:.....:
1941 AAGAAGAAAGCGGACGCCCTATTGCTGGCGCATTTGCTTAATTGA 1990
353 pProIleValGlnGluSerSerLeuGluSerProLeuProSerSerL 370
:|||||:.....: |||||||:.....: |||||:.....:
1991 C.....GCCATA 1998
370 euGlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 386
:|||||:.....: |||||||:.....: |||||:.....:
1999 TTAAAGCTCAAGTAGTGTGATTCGCTAAACCGGTGT.....TTT 2042
387 IleGlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSe 403
:|||||:.....: |||||||:.....: |||||:.....:
2043 GTGAGACATCGGGCATTTATTCATTCAGACCAATGCAATGTGTA 2092
403 rGluGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnG 418
:|||||:.....: |||||||:.....: |||||:.....:
2093 AACAAAGAGTGTGTTAGACCTGATGATGTAACATTGAACCGGAG 2142
418 InProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
:|||||:.....: |||||||:.....: |||||:.....:
2143 ACCCCTTGCAATATATACCGT.....ATPAATGATGAATTCCTCA 2183
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAl 446
:|||||:.....: |||||||:.....: |||||:.....:
2184 ACAGCAGCGGTGAAGCAAGCAGCACCTTAAAAAAATACGAACTCAAAAC 2233
446 aLeuLeuIle..... 449
:|||||:.....: |||||||:.....: |||||:.....:
2234 AACGCTAACCAATACAACTATTTCAAATTATCTGAAAAACCGCTGACAA 2283
450 .....MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
:|||||:.....: |||||||:.....: |||||:.....:
2284 TGAATATACCGCATCAAGAAACTTACCGTTAATAGCTCAATCAACATC 2333
465 AlaThrLeuSer.....IleProLeuHisSer..... 473
:|||||:.....: |||||||:.....: |||||:.....:
2334 GGAACCACTCCCACTTAATCTCCATAGTAAAGCTCAGCGTGGCGGAG 2383
473 ..... 473
2384 GGTTCAGATTGAGATATTACTTTAAAGCGCAATTTAACCATTT 2433
474 .....LeuAspThrGluLysSerValThrIleHisAlaPro 485
:|||||:.....: |||||||:.....: |||||:.....:
2434 ATTCGCGCGATGCGTTGATGTTCTATAAAATTTACCGTTGATCAGGCT 2483
486 AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPh 502
:|||||:.....: |||||||:.....: |||||:.....:
2484 TTTTAAATATTACCGCGCGCTTCCTTGTGAGAGTGCAGAAATACAA 2533
502 eTyrGlu.....AsnValGluLeuLeuSerLys..... 511
:|||||:.....: |||||||:.....: |||||:.....:
2534 AGCAGCGACGCGCAATGCTAAATTTGCGCCAGCGCACTGAACA 2583
512 .....GluGlnAsnAsnIleProLeu..... 518
:|||||:.....: |||||||:.....: |||||:.....:
2584 TTACAGCAGAGCAAAACATTTCAAGGCTACACACAGCTATCTTAAACGA 2633
519 .....LeuThrLeuSerLysGlnGlnSerHisLeuHisLeuPr 531
:|||||:.....: |||||||:.....: |||||:.....:
2634 ACGGTAAGGCTCGAATATATCTTTCATCAGCATTAATTTA..... 2675
531 aspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPhePheS 548
:|||||:.....: |||||||:.....: |||||:.....:
2676 ACCCACAATCTTACTGACAAATTAACATATCTGGATATTAACAATTA 2724
548 eF.....TrpLysAspSerAspGluGly 555
:|||||:.....: |||||||:.....: |||||:.....:
2725 ACCAACTACGAAAGAAAGAACCTCTGTTTGGCAACCGACCATGATTG 2774
556 HisSerLeuIle.....AlaAsnTrpThrPr 564
:|||||:.....: |||||||:.....: |||||:.....:
2775 CACTGGAAGCTGAGTCGCTTAATCTAGACAGCGCAATTTTACTCT 2824
564 OlysAsnTyrValProHisProGluArgGlnSerThr..... 576
:|||||:.....: |||||||:.....: |||||:.....:
2825 TATTAATATACATTTCAGCAATAGCAAGGCTTAACAACACAGATAGAA 2874
576 ..... 576
:|||||:.....: |||||||:.....: |||||:.....:
2875 GCTTGACAGGGTGAATTTTAAAGCGCTAAATGCCACATGCTCATTCAT 2924
577 .....LeuValAlaAsnThrLeuTr 583
:|||||:.....: |||||||:.....: |||||:.....:
2925 CTCGAAGAAGGAGCGAAATTTCAATTTCAAAATTAACCAACGACACAT 2974
583 pasnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleA 600
:|||||:.....: |||||||:.....: |||||:.....:
2975 GAACACA...AGCAAACTTTACCAATTCGTTTTCACCAATATACAG 3021
```

```

1 REFERENCE/DOCKET NUMBER: 1038-404
2 TELECOMMUNICATION INFORMATION:
3 TELEPHONE: (703) 415-0810
4 TELEFAX: (703) 415-0813
5 INFORMATION FOR SEQ. ID NO.: 6:
6 SEQUENCE CHARACTERISTICS:
7 LENGTH: 9133 base pairs
8 TYPE: nucleic acid
9 STRANDEDNESS: single
10 TOPOLOGY: linear
11 MOLECULE TYPE: DNA (genomic)
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
```

```
alignment_scores:
  Quality: 148.50
  Ratio: 0.408
  Percent Similarity: 44.828
  Length: 812
  Gaps: 36
  Percent Identity: 18.227
```

```
alignment_block:
US-09-677-752-4 x US-08-302-832-6
```

Align seg 1/1 to: US-08-302-832-6 from: 1 to: 9323

[illegible][illegible]


```

577 .....LeuValAlaasnThrLeuTrpAs 584
2488 AAGAAGGAGCGAAGTAAATTCAATTAAACCAAGACGACACTGA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleasnThrIleAla 601
2538 CACA...ACCAAACTTTACCAATTCGGCTTTTACCATATACACGCCA 2584
601 IsGlyGlyValTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
2585 CTGGTCG.....GGCTCTGTTTTTTTTCATATA 2613
618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnThrPheIst 633
2614 TATGCCAACCATCTGCGCAGAGGGCTGATTAAATAGTGAATTAA 2663
633 sarGSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsp 650
2664 TATCTCTAACGGCGGCTTAATTTCACCTTAATCCCATGTTCCCGCGAG 2713
650 IsSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACCCTTTTAAATTC.....AACAAAGACTTAA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
2740 ACCATAATGCACCAATCA..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlu 700
2761 .....AATTTC 2767
700 eTlleHisGluLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCGTCAGACGACGAAAGATGATTTTATGACGGCTACCGCAGCAATGCC 2817
716 PheGlySerThrPheIleSerValAlaValSerGly 726
2818 ATCATTCACCTACACATATCATCTCTGGCG 2850

seq_name: /cgn2_6/prodata/2/ina/58_COMB.seq:US-08-617-697-3
seq_documentation_block:
: Sequence 3, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-APR-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166

```

```

FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
: NAME: Berksstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4937 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-617-697-3

alignment_scores:
Quality: 148.50 Length: 811
Ratio: 0.408 Gaps: 35
Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-617-697-3
Align seg 1/1 to: US-08-617-697-3 From: 1 to: 4937

73 AsnSerTyrCysTrpPheValSerLysLeuHisIleThrAspProLysG 89
778 AAGCGACAAGCTCTTTTATATCAACCCAAATGCTATACATAAGTAAGA 827
89 uAlaLeuPheLysGlyLysGly.....AspLeuSerI 100
828 CCGCAATTATTACATAATAGCGCTTACGGCTTACGCTAGCAATTCTCA 877
100 LeuGlnAspPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116
878 ACGAAACATCAAGCGCGTAAATTCACCTTCGACCAACCAAAAGATAA 927
117 SerProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAs 133
928 GCGCTCGCTGAATTT...GTGAATCAGCGTTTAATTACTGTCGTAAGA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluLysGlyGlyAla 150
975 CGGCAGTGTAATCTTATT.....GGTGCAGAG 1003
150 LeuSerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
1004 TGAAA..... 1008
167 GluGlnAsnSerSerLysGlyAsnGlyAlaIleGln.....Al 180
1009 AAGCAGCGTGATGATTAGCGTAATGCTGCGACGATTTCTTACTCGCAG 1058
180 aGlnThrPheSerLeuSerArgAsnValSerPro.....I 192
1059 GCAAAATAATCACACATCACGATATATTAACCAACCATTAATTACACA 1108
192 IsSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleCysGlyS 208
1109 TTGCGCGCGCTGAATAATGAGCG...GTCAATCTGGGGGATATTTTGGC 1155
209 SerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThrGlyAs 225
1156 AAA.....CGGCTAATATATGTC.....CG 1178
225 nSerAlaThrAsnGlyAlaAlaIleCysCysIleSerAspLeuAsnThrS 242
1179 TGCTGCCACT..... 1188
242 eGlyLysGlySerLeuSerLeuAlaCysAsnGlnGlnIuThrLeuPheAla 258

```

2047 TTAATATTACCGCCGCTTCCGTAGCTTTTGAAGTGGAAATACAAAC 2096
503 :glu.....AsnValGluLeuLeuSerLys..... 511
2097 ACGCAGCGGGGAAATGCTAAATTTGTCCGCCAGGGGCACTGTAACCTTA 2146
512 :.....GluGlnAsnAsnIleProLeu..... 518
2147 CAGGAGAGGAAAGATTTCAGAGGCTTACACGCTATCTTTAAAGGAAAG 2196
519 :.....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAs 532
2197 GGTAAAGCTCTGAATATATCATTCATCAGTAAATTTA.....AC 2237
532 pCysAsnLeuSerSerHisPheGlyTyrGlnGlyAspTyrPheSer 548
2238 CCACATCTTACTGGCAGACATTAACATATCTGGAAATATACAAATTAAAC 2287
549 :.....TyrLysAspSerAspGluGlyHis 556
2288 AACTACAGAGAAAGAACCTCGTATTTGGCAACACGACATGATTCGCAC 2337
557 SerLeuIle.....AlaAsnTyrThrProLys 565
2338 TCGAAGCTCAGTCTCTTAATCTAGACAGCGGCAAAATTTTACCTTTAT 2387
565 sAsnTyrValProHisProGluArgGlnSerThr..... 576
2388 TAAATATACATTTACAGCAATAGCAAAAGGCTTACACACAGTATAGAGCT 2437
576 :..... 576
2438 CTGACGGGTGAATTTTAAAGCGGTAAATGGCAACATGTCATTAATCTC 2487
577 :.....LeuValAlaAsnThrLeuTrpAs 584
2488 AAGACAGACGCAAGTTAATTCAATTTAAACCAACAGCAACATGCA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetLeuAsnThrIleAla 601
2538 CACA...AGCAAACTTTACCAATTCGGTTTACCAATATACAGCA 2584
601 isGlyValAlaTyrLeuPheGlyThrTyrGlySerAlaValSerAsnLeu 617
2585 CTGGTGG.....GGCTCTGTTTGTGTTTGTATATA 2613
618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTyrHisI 633
2614 TATGCCAACCATTTCTGGCAGAGGGGCTGACITTAATAATCAGTGAATTA 2663
633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsp 650
2664 TATCTTAACGGCGTAATTTTACCTTAATTCCTTCCATGTCGGCGCATG 2713
650 isSerPheCysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerAspSer 666
2714 ACGCTTTTAAATC.....ACCAAGACCTTA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAla 683
2740 ACCATTAATGCAACCAATTCA..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlu 700
2761 :.....AATTCA 2767
700 eTlleHisIleLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCGTCAGACAGCAAGATGATTTTATAGCGGGTACCGCAATGACC 2817
716 PheGlySerTyrPheIleSerValAlaValSerGly 726
2818 ATCAATTCACCTACACATATCATCTTCGGC 2850

seq_name: /cgn2_6/prodata/2/ina/58_COMB.seq:us-08-728-470-3
seq_documentation_block:
: Sequence 3, Application US/08728470
: Patent No. 5928651
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/728,470
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berksstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-633
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0810
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4937 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-728-470-3

alignment_scores:
Quality: 148.50 Length: 811
Ratio: 0.408 Gaps: 35
Percent Similarity: 44.883 Percent Identity: 17.879

alignment_block:
US-09-677-752-4 x US-08-728-470-3 ..
Align seg 1/1 to: US-08-728-470-3 from: 1 to: 4937

73 AsnSerTyrCysTyrPheValSerLysLeuHisIleThrAspProLysG 89
||||| :|||||
778 AAGCGACAGGCTTTTATATCAACCAATGCTATACATAGGTAAGCA 827
||||| :|||||
89 uAlaLeuPheLysGluLysGly.....AspLeuSerI 100
:||||| :|||||
828 CCAATTTATATACATAATGCGCTTACGGCTCTACGCGAGACATTTCA 877
||||| :|||||
100 IecIAsnPheArgPheLeuSerPheThrAspCysSerSerLysGluSer 116

```

1656 AAAAGAGTGCTGTAGACCTGATGATGATACATTTGACGCCGAAGACC 1705
      |||||.....|
419 IolIeGluLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
      |||||.....|
1706 CCTTCGCATATATACCGT.....ATATGATGATTTCCCAACA 1746
433 .....AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaIle 447
1747 GGCACCGGAGCAAGCAGCCCTAAAAAATACGCACTCAAAACAC 1796
447 uLeuIle..... 449
1797 GCTAACCATACACTATTTCATCTGAAAAACGCTGCACATCA 1846
450 ..MetGluAlaGlyThrSerLeuLysThrSerAspLeuLysLeuAla 465
1847 ATATACCGCATCAAGAAACTTACCGTTATATAGCTCATCAATCAATCGGA 1896
466 ThrLeuSer..IleProLeuHisSer..... 473
1897 AGCAACTCCCACTTATCTCATAGTAAAGTCAAGCTGCCGAGCGGT 1946
473 ..... 473
1947 TCAGATGATGAGATATTACTCTAAAGCGGCAATTTAACCATTTATT 1996
474 .....LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
1997 CTGCGGATGGGTGATGATGCTTCAAAAAATATACCTGATCAGCGTTT 2046
487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAspPheLys 503
2047 TTAAATATTACCGCGCTTCCGCTACCTTTTGAAGGTGCAAAATCAAAAC 2096
503 rGlu.....AsnValGluLeuLeuSerLys..... 511
2097 ACGCAGCGGCAATGCTAAATGTGCGCCGACGCACTGAACCATTA 2146
512 .....GluGlnAsnAsnIleProLeu..... 518
2147 CAGGAGAGGAAAGATTTCAGGCGCTAACACGATCTTTAAACGGAACG 2196
519 .....LeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAs 532
2197 GGTAAAGCTCTGAATATCATTTTCATCAGTAATTA.....AC 2237
532 polyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrProPheSer. 548
2238 CCAGATCTTAGTGGCACATTAATCATATCTGGATATATACATTAAC 2287
549 .....TrrPlyAspSerAspGluGlyHis 556
2288 AAACCTACGAGAAAGAACACCTGTATGGCAAAACGACCATGATTCGCAC 2337
557 SerLeuIle.....AlaAsnTrpThrProLys 565
2338 TGGAACGTCAGTGCCTTAATATAGACAGCGGCAAAATTTTACCTTAT 2387
565 sAsnTyrValProHisProGluArgLysThr..... 576
2388 TAAATACATTTCAAGCAATAGCAAAAGGCTTAACAACACAGTATAGACT 2437
576 ..... 576
2438 CTGACGGGGTGAATTTTAACGGCGTAAATGGCAACATGTCATTCAATCTC 2487
577 .....LeuValAlaAsnThrLeuTrpAs 584
2488 AAAGAGAGCGAAAGTTAATTTCAATTTAAACCAACGAGACATGAA 2537
584 nThrTyrSerAspMetGlnAlaValGlnSerMetLeuAsnThrIleAlaIle 601
      |||||.....|

```

```

2538 CACA...ACCAACTTTTCCCAATTCGGTTTTTACCCAATATACAGCCA 2584
601 isGlyAlaValTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
2585 CAGTGGC.....GGCTCTCTTTTTCATATTA 2613
618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTrpHisHis 633
2614 TATGCCAACCAATTCGCGACACCGGCTGAGTTAAATGAGTGAATTA 2663
633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsn 650
2664 TATCTCTAACGGCGCTAATTTTACCTTAATTCCTGATGCTGGCGGATG 2713
650 isSerPheCysLeuAlaIaGlyGlnLeuGlyLysSerSerAspSer 666
2714 ACGCTTTTAAATC.....AACAAACCTTA 2739
667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
2740 ACCATTAATGCAACCAATTCA..... 2760
683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGlu 700
2761 ..... 700
700 eTlleHisGluLeuLysThrLys...TyrArgSerPheSerLysGluGly 715
2768 GCGTCAGACAGCAGCAACATGATTTTATGACGGTACGACGCAATGCC 2817
716 PheGlySerTrpHisSerValAlaValSerGly 726
2818 ATCAATTCAACCTACACATATTCATCTCGGC 2850

seq_name: /cgn2_6/prodata/2/ina/5b_comb.seq:us-08-469-880-3
seq documentation block:
; Sequence 3, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; City: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651

```

259 SerAsnSerAlaLysGluLysGlyAlaIleTyr.....AlaLysHis 273
 1216 GATTCTGTAAGCAAGATAAAGCGCAATATTGTTCTTTCGCCAAGA 1265
 273 sMetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAla 290
 1266 GCGTAAGCGGAATATGGCGGTGTATTTCCGCTCAAAATCAGCAGCTA 1315
 290 ys.....IleGlyGlyAlaIleAlaIleGlnSerGly 300
 1316 AAGCGCGCAACCTGATGATTACAGCGCATTAAGTCACATTAAACAGCT 1365
 301 GlySerLeuSerIleLeuAlaGlyLysSerValLeuPheGlnAsn 317
 1366 GCAGTTATCGACCTTTCAGGTAAAGAGGAGCAAACTTACCTTCGCCG 1415
 317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu 334
 1416 TCACGAGCC.....GGCGAAGGTAAACGGCATTCATTAATG 1453
 334 LuLysAspAlaIleLeuSerSerLeuGluAlaArgAsn..... 346
 1454 CAAGCAAAACCTCTTTAGAAAAAGCTCACCATCATGATATCAGGCAAA 1503
 347GlyAspIleLeuPhePheAspPr 354
 1504 GAAAAAGCGGAGCGCCATTTGTCGCCGCGATTTGCCCTTAATTCAC... 1551
 354 oIleValGlnIleSerSerSerLysGlySerProLeuProSerSerLeuG 371
 1552GGCAATATTA 1561
 371 InAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
 1562 ACCCTCAAGGTAGTGATATCGCTAAACCGGTGT.....TTTGG 1605
 388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
 1606 GAGACATCGCGCATTTATTCATTCAGACCAATTCATTCATTAAC 1655
 404 uGluGlu.....LysThrProAspAsnLeuThrSerGlnLeuGlnIle 419
 1656 AAAGAGGTGTTGCTAGACCTGATGATGATTAACATTGAGCGGAGAC 1705
 419 rIleGlnLeuLysSerGlyArgLeuValLeuLysAspArg..... 432
 1706 CCTTCGCAATATACCGT.....ATTAATGATGAATTCACACA 1746
 433AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAla 447
 1747 GGCACCGGTGAAGCAAGCGACCTTAATAAATAGCAACTCAAAACAC 1796
 447 uLeuIle..... 449
 1797 GCTAACCAATACAACTATTCTCAATATCTGAAAAACCGCTGGACAAATGA 1846
 450 .MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
 1847 ATATACCGCATCAAGAAACTTACCTTAATAGCTCATCAACATCGCA 1896
 466 ThrLeuSer...IleProLeuHisSer..... 473
 1897 ACCAATCTCCACTTAATTCATAGTAAGGTCAGCGTGGCGAGCGCT 1946
 473 473
 1947 TCAGATGATGAGATATTACTCTTAAGCGGAATTTAACATTATT 1996
 474LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
 1997 CTGCGGATGGTGGTGTATGATTAATAAATATTACCTTCAGACCGGTTT 2046
 487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503

2047 TTAATATATACCGCGCTTCCTACCTTTTGAAGGTGAATAATACAAAGC 2096
 503 rGlu.....AsnValGluLeuLeuSerLys..... 511
 2097 ACGCAGCGCGCAAAATGCTAAATATGTCCGCGACGACATGTAACATTA 2146
 512GluGlnAsnAsnIleProLeu..... 518
 2147 CAGGAGAGGAAACATTTACAGCGCTAACACATCTTTTAAACGGAACG 2196
 519LeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAs 532
 2197 GGTAAAGCTGTGAATATCATTTTCATCAGTGAATTA.....AC 2237
 532 pGlyAsnLeuSerSerHisPheGlyTyrGlnGlnLysPhePheSer 548
 2238 CCACATCTTACTGTGCACAAATTACATATCTGGGAATATACAAATTAAC 2287
 549TrpLysAspSerAspGlnGlnHis 556
 2288 AACCTACGAGAAAGAACACCTGCTATTGGCAACCGCATGATTCGCAC 2337
 557 SerLeuIle.....AlaAsnTrpThrProLys 565
 2338 TGGAGCTGAGTGTCTTAATCTAGACAGCGCAAAATTTACCTTAT 2387
 565 sAspTyrValProHisProGluArgGlnSerThr..... 576
 2388 TAAATACATTTCAAGCAATAGCAAAAGCTTAACAAACAGATATAGAGCT 2437
 576 576
 2438 CTGCAAGGGTGAATTTTAACGGCGCTAAATGCGACATGTCATTCATCTC 2487
 577LeuValAlaAsnThrLeuTrpAs 584
 2488 AAAGAGGAGCGGAAGTAAATTTCAATTAACCAACGAGAACATGAA 2537
 584 nThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 601
 2538 CACA...AGCAACCTTTACCAATTCGGTGTATTTACCAATATCAGAGCA 2584
 601 IsGlyAlaLysIleuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
 2585 CTGCTGGC.....GCCTCTTTTTCATATTA 2613
 618 PheTyrAlaHisAspSerSerGlyLysProIleAsp...AsnTrpHisI 633
 2614 TATGCCAACCATTTCTGCGCAGAGGCGCTGATTAAATAGAGTAATTA 2663
 633 sArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsn 650
 2664 TATCTTACGCGCTAATTTTACCTTAATTCCTATTCATTTGCGGCGATG 2713
 650 IsSerPheCysLeuAlaIleGlnLeuLeuGlyLysSerSerAspSer 666
 2714 ACCGTTTAAATC.....ACCAAGACCTTA 2739
 667 PheIleThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaG 683
 2740 ACCATTAATGCAACCATTA..... 2760
 683 nLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrAsnGln 700
 2761AATTTC 2767
 700 rIleHisGlnLeuLysThrLys...TyrArgSerPheSerLysGlnGly 715
 2768 CCGTCAGACAGCAAGATGATGATTTTATGACGGGTACGACGCAATGCC 2817
 716 PheGlySerTrpHisSerValAlaValSerGly 726

```
117 SerProSerIleIleHisGlnLysAsnGlnLeuSerLeuArgAsnAs 133
    ::::: ||| ::::: ||| ::::: |||
928 GCGCTCGCTGAATTT...GTGAATCAGCGTTTAATTAATGTCGCTAACA 974
133 nGlySerMetSerPheCysArgAsnHisAlaGluGlySerGlyAlaI 150
    ||||| ::::: |||
975 CGCGACTGTAATCTTATT...CGGGCCAAG 1003
150 leSerAlaSpAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPhe 166
    :::::
1004 TCAAA... 1008
167 GluGluAsnSerSerLysGlnGlyAlaIleGln...Al 180
    ::::: ||| ::::: ||| ::::: |||
1009 AACGAGGCGTGTATGCTTAATGCTGACGATTTCTTACTCCGACG 1058
180 acGlnThrPheSerLeuSerArgAsnValSerPro...T 192
    ||| ::::: ||| ::::: |||
1059 GCAAAAAACACCATCAGCATATATAAACCCACCATTAATCTTACACCA 1108
192 leSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysCys 208
    ||| ::::: ||| ::::: |||
1109 TTGCGCGCGCTGAATATGACG...GTCAATCTGGCGCATTTTGTGCC 1155
209 SerAsnLeuIleCysSerClyAsnValAsnProLeuPhePheThrClyAs 225
    ::::: ||| ::::: |||
1156 AAA...GGCGTAACATTAATGTC...CG 1178
225 nSerIleThrAsnGlyGlyAlaIleCysCysIleSerAspLeuAsnThrS 242
    ::::: |||
1179 TCCTCCCACT... 1188
242 erGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 258
    ||||| ||||| |||
1189 ...ATTCAACCAAGTAAGTAACCTTCTGCT 1215
259 SerAsnSerAlaLysGlnLysGlyAlaIleTyr...AlaLysHis 273
    ::::: ||| ::::: |||
1216 GATCTGTAAAGCAAGATTAAGCGCAATATTGTTCTTCGCCAAGA 1265
273 smetValLeuArgTyrAsnGlyProValSerPheIleAsnAsnSerAlaL 290
    ::::: ||| ::::: |||
1266 GCGTCAACCGCAATTTGCCGCTGTAATTTCCCTCAAAATCAACAAGCTA 1315
290 ys...IleGlyGlyAlaIleAlaIleGlnSerGly 300
    |||
1316 AAGCGGCAAGCTGATGATTACAGCGCATTAAGTACATTAATAACAGCT 1365
301 GlySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGlnAsnAs 317
    ::::: ||| ::::: |||
1366 GCAGTTATCGACCTTTCAGTAAAGAGGCGAAGCAACTTACCTTGGCGG 1415
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
    ||| ::::: |||
1416 TCACAGCGCC...GGCGAAGTAAATAACGCAATTCATTAG 1453
334 LuLysAspAlaIleLeuSerSerLeuGlnAlaArgAsn... 346
    ||| ::::: |||
1454 CAAGAAACCTCTTTAGAAAAAGGCTCAACCATCAATGTATCAGGCAAA 1503
347 ...GlyAspIleLeuPhePheAspPr 354
    ||||| ::::: |||
1504 GAAAGGCGACGCGCTATTGTGTGGCGCATTTGCGTTAATGAC... 1551
354 oileValGlnGluSerSerLysGlnSerProLeuProSerSerLeuG 371
    ::::: |||
1552 ...GCCATATATA 1561
371 LuAlaSerValThrSerProThrProAlaThrAlaSerProLeuValIle 387
    ::||| ::::: |||
1562 ACGCTCAAGTAGTGGTATCGTAAACCGGTGT...TTTGTG 1605
388 GlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
```

```
1606 GAGACATCGGGCGATTAATTATCCATTGACAGCAATGCAATGTTAAAC 1655
404 uGluGlu...LysThrProAspAsnLeuThrSerGlnLeuGlnP 419
    ::||| ::||| ::|||
1656 AAAAGAGTGCTGCTAGACCCGATGATGATTAACAATGAAGCGAAGAC 1705
419 rolleGluLeuLysSerGlyArgLeuValLeuLysAspArg... 432
    ||| ::||| ::|||
1706 CCGTTCGCAATATACCGGT...ATAAATGATGAATTCGCCACA 1746
433 ...AlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaIle 447
    ::||| ||| ::|||
1747 GGCACCGCGTGAAGCAAGCACCCCTAAAAAATAGGAACCTCAAAACAC 1796
447 uLeuIle... 449
    |||
1797 GCTAACCAATACACTATTTCAAATATCTGAATAAACGCCCTGACAATGA 1846
450 ...MetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAla 465
    ::||| ::||| ::|||
1847 ATATAACGGCATCAAGAAACCTTACCGTTAATAGCTCAATCAACATCGGA 1896
466 ThrLeuSer...IleProLeuHisSer... 473
    ::||| ::|||
1897 ACCAATCCCACTTAATCTCCATAGTAAGGTCAAGCTGCGGAGCGCT 1946
473 ... 473
1947 TCAGATTGATGAGATATTACTTCTAAAGCCGCAATTTAACCATTTATT 1996
474 ...LeuAspThrGluLysSerValThrIleHisAlaProAsn 486
    ::||| ::||| ::|||
1997 CTGGCGGATGGGTGATGCTTCAATAAAATATTCCTGATCAGGGATT 2046
487 LeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspGluAsnPheTy 503
    ::||| ::|||
2047 TTAATATTACCGCGCGCTCCCTACCTTTTGAAGGTGCAATAAACAAGC 2096
503 rGlu...AsnValGluLeuLeuSerLys... 511
    ::|||
2097 ACCGACGCGGCAAAATGCTAAATTTGCGCCAGCGCACTGTAACCATTA 2146
512 ...GluGlnAsnAsnIleProLeu... 518
    |||
2147 CAGGAGAGGAAAAAGATTTCAGGGCTTAACAACGTAATCTTAAACGAAAG 2196
519 ...LeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAs 532
    ||||| ::|||
2197 GGTAAAGCTGATATATCATTTTCATCAGTGAATATTTA...AC 2237
532 pGlyAsnLeuSerSerHisPheGlyTyrGlnGlyAspThrPheSer 548
    ||||| ::|||
2238 CCACATCTTACTGTGACAAATTAACATATCTGCGAATATACAAATTAAC 2287
549 ...TrpLysAspSerAspGlnGlyHis 556
    |||||
2288 AAATCAGCAAGAAACACCTCGATTGTCGAACCCAGCATGATTCGAC 2337
557 SerLeuIle...AlaAsnTrpThrProLys 565
    ::|||
2338 TGGAACGTCGATGCTTAATCTAGACAGACGCCCAATTTTACCTTTAT 2387
565 sAsnTyrValProHisProGluArgGlnSerThr... 576
    ::||| ::|||
2388 TAAATACATTTCAAGCAATACCAAGGCTTAACAACAGATAGAAAGCT 2437
576 ... 576
2438 CTGCAAGGCTAATTTTACGCGCTAAATGCAACATGCTTTCATTAATCTC 2487
577 ...LeuValAlaAsnThrLeuTyrPAs 584
    ||| |||
```

```
784 yllethphcjlulysyssercin...lysthrargfthrtyrtyr 800
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23567 GATCCGATTGTGATAAGCAATCAGACTGCCAAGATGCACGTACATCTAA 23616
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
800 heleuglyalaTyrIleGlnaspLeuLysarGaspValGluSerGlyPro 816
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23617 CTCCTTGCT...TATACGTGTGATCTTGTCTGTACTAACCCCGACTGTACG 23663
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
817 ValValleuLeuLysAsnAlaValSerTrpAspAlaPrometAlaAsnLe 833
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23664 ACAACACTCGCAATTACCGCTGATTTGCAAAACTTTCGTACGAATTT 23713
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
833 uaspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgAlaLeuHis. 849
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23714 GGCAGAGACAAGCTTTAGCTCCTCGTCGAGGAAACCAATTTTGTCTTA 23763
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
850 . ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSer 865
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23764 CAATTTTGAAGCCTTAGCCCAATTTTCTTTTCAATTGCGTGGCTCATCT 23813
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
866 HisSerTyrSerLeuAspLeuGlyThrThrTyrArgPhe 878
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
23814 CGCAATTACATGTAGACTTAGAGCAAAATACCAATTC 23852
```



```
Align seg 1/1 to reverse of: AE002235 from: 1 to: 12127

126 GlycylLeuSerLeuArgAsnAsnGlySerMetSerPheCysArgAsnI 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8673 GGGTAGCTGTGAGTTGACAAAATGTCAGTTGCTCTTACACAAAACCTT 8624
142 sAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuGln 159
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8623 TTCAACGAGTATGCGCTCTATCACCCCAAAAACCTCTTCATTAACAG 8574
159 IsAsnTyLeuPheThrAlaPheGluGluAsnSerSerLeuGlyAsnGly 175
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8573 GGACTACATCTCAGCTCTGTTTCTGAAAATACCTCTCTCAAGAAAGCG 8524
176 GlyAlaIleGln...AlaGlnThrPheSerLeuSerArgAsnValSerPr 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8523 GGACCCATTCAGACTCCGATGCCCTTACCTTACTGAAAACCAAGGGA 8474
191 oIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAlaIleCysC 208
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8473 AGTCCTTTTCTGCACAACTACTCTTCGATTTCTGACCTGCAATTTTGA 8424
208 ySSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePheThyrGly 224
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8423 CAGAAAGCTCGTGACTATTCTAATAATGCTAAAGTTCTCTTATTGAC 8374
225 AsnSerAlaThr.....AsnGlyC 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8373 AATTAAGTCACAGAGAGAGCTCCTCAACAACGGGGATATGTCAGAGAG 8324
231 yAlaIleCysCysIleSerAspLeuAsnThrSerGlyGlySerLeuS 248
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8323 TGCATCTCTGCTCT.....TATAAACACACATACATCTTAACGCA 8283
248 eLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaGlyGlu 264
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8282 CCTCCACGGAATCAATGTTACTCTTACACCAACAATACATCAACAACA 8233
265 LysGlyGlyAlaIleTyAlaAlaHisMetValLeuArgTyArgAsnGlyPr 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8232 GCGGAGGAGCTATCTATGTGAAAAGCTGCAACTGCGCTCCGAGAGACT 8183
281 oValSerPheIleAsnAsnSer.....AlaLysIleGlyG 293
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8182 TACCTTATTCAGTACAAATACGTCAATGACAGTACACCTCTTAACGTC 8133
293 LyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGlu 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8132 GAGCCATAGCTATCGAAGATAGTGGGGAATTGATTCGCCGATAGT 8083
310 GlySerValLeuPheGlnAsnAsnSerGlnArgThrSerAspGlnGlyLe 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
8082 GGTGACATTTGCTTTTAGGGAATACAGTCACTTACTACTCCTGGGAC 8033
326 uValArgAsnAlaIleTyArgLeuGlyLysAspAlaIleLeuSerSerLeuG 343
    : : : : : : : : : : : : : : : : : : : : : : : : : :
8032 GAATGAGAGTATGATCGACTTACGACAGACGACGCAAGATGACACTTTC 7983
343 LuAlaArgAsnGly...AspIleLeuPhePheAspProIleValGlnGlu 358
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7982 GTTGTGCTGTGTAGAGCCATCTACTTATATGATCCATAACTACAGCA 7933
359 SerSerSer.....LysGluSerProLeuPr 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7932 TCATCCACACACAGTACAGATGCTTTAAAGTAAATGAGACTCCGGCAGA 7883
367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7882 TTCTGCACATACAAATATACA..... 7864
384 roleuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGlu 400
    : : : : : : : : : : : : : : : : : : : : : : : : : :

7863 .....GGGAACATCATCTTCACAGAGAA 7840
401 ArgLeuSerGluGluGlyLysThrProAsp...AsnLeuThrSerGln 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7839 AAGTTATACAGACAGAGAGCCGACATTTCTAAATCTTACTTACGAACT 7790
416 uGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAspArg 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7789 ACTACAGCTCTTAACCTTTTTCAGAGAGTACTTACTTAAACATGAG 7740
433 lValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIle 449
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7739 TGACTCTGCAAGCTCAGGACTTCACTCAACAGCAGATTCCTGCTCGAA 7690
450 MetGluAlaGlyThrSerLeuLysThrSerSerAspIleuLysLeuAla 466
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7689 ATGACAGCTAGAACTACTCTAGAACCTCTGATACTACACCATTAACA 7640
466 lLeuSerIlePheProLeuHisSerLeuAspThrGlyLysSerValThrIle 483
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7639 TTTCGTCATTAACATCACTTCTATAGACGTCGCAAGAGCAAAATAG 7590
483 lS.....AlaProAsnLeuSerIleGlnLysIlePheLeuSer 495
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7589 AAACCAAGCTAGCTCAAAAATCTGACTTATCTGGAACCATCACTTGA 7540
496 AsnSerGlyAspGluAsnPheTyArgAsnValGluLeuLeuSerLysG 512
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7539 TTTCGACCCGAGCGGCACCTTTTATGAAATCATAGTTTAAACATCCCA 7490
512 uGlnAsnAsnIleProLeuLeuThrLeuSer...LysGluGlnSerHisL 528
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7489 GTCCACACAGCTTACAGAGCTCAAAAGCTTTCGAACTGATACAGACCG 7440
528 euHisLeuProAspGlyAsnLeuSerSer.....HisPheGlyTyArgGln 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7439 CAGTACTCCAGATCTATATATGGTGAGAAATTCATCTAGCGGTATCAG 7390
543 GlyAspThr...ThrPheSerTriPlysAspSerAspGluGlyHisSerLe 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7389 GGAAGCTTGGGCCCAATTTGTTGC.....CGACAGGCGCTTCTAC 7349
558 uIleAla.....AsnTriPThrProLysAsnTyValProHisProGlu 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7348 GACTGCAACCTTCAACCTGACTAAACCTGCTATATCTTATCCCGACG 7299
573 rGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTySerAspMet 589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7298 GTATCGGCTCTTATGCCCTATATAGCTTATGAAATGCTTATAGATAT 7249
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGly.....GlyAl 604
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7248 AGCTCTCCCATTTATCTTATATGAGAGCTGCAACGAGGTTCCAGCGAGA 7199
604 aTyLeuPheGlyThrTriPglySerAlaValSerAsnLeuPheTyArgHis 621
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7198 CCGTGCTTTT.....TGGTGTCTGATATCTTAACCTTCTTC.....C 7161
621 lAspSerSerGlyLysProIleAspAsnTriPHisHisArgSerLeuGly 637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7160 ATAGAGATAGTACAAAAACACAGCGCGGTTTCGCATTTGAGCGCGT 7111
638 TyrLeuPheGlyLysSerThrHisSerLeuAspHisSerPheCysLe 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7110 TATGTCATAGAGAGAAACCTACATCTTGTTCAGATTAAGATTTCTTAGTC 7061
654 uAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerPheIleThrSer 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7060 TGCATTTTGTACGCTTTTGAAGAGATAGAGACTACTTCTTACCTAGA 7011
671 hrcGluThrThrSerTyrlleAlaThrValGlnAlaGlnLeuAlaThrSer 687
    : : : : : : : : : : : : : : : : : : : : : : : : : :
7010 ATCAAGGTACAGTCTACGAGAGAACTCTTATCTTACAGCAACAGAAACC 6961
```

```
3524 CAGCTTATGTTAGATTCCTATCTAACATAGCGTCTACCTCAGAGCGCCT 3573
206 IleCysSerSerLeuIleCysSerGlyValAsnProLeuPhePhe 222
||| : : : : : ||| : : : : :
3574 ATCGATGATGAGGACGCTCGATCTACTATCGAACACAAATTTCTATATT 3623
222 ethrGlyAsnSerAla...ThrasnglyAlaIleCysCysIleSera 238
| | | | | : | | | | | : | | | | |
3624 TGAAGGGAAATCGACGAAACTACTGGCGGTGCATCTGC... 3663
238 sPLeuAsnThrSerGlySer...LeuSerLeuAlaCysAsn 252
| | | | | : | | | | | : | | | | |
3664 ....AACACCAAGCGAGCTGATCTCTGAACTGATATCTCTAACAAAT 3708
253 GlnGluThrLeuPheAlaSerAsnSerAlaLysGlyValGlyAlaIle 269
||| : | | | | | | | | | | | : | | | | |
3709 AAGACTCGATCTTGGCTTCAACGTAGCAGAACAAACGGGTGGCCCAT 3758
269 eTyraAlaLysHisMetValLeuArgTyrAsnGlyProValSerPheIle 286
| | | | | : | | | | | : | | | | |
3759 CCATGCTTAAAGCTAGCCCTTCTCTGAGGCTTACAGAGTTTCTAC 3808
286 snAsnSerAlaLysIle...GlyGlyAlaIleAlaIleGln 298
||| : | | | | | | | | | | | : | | | | |
3809 GAAATTAATGTCATCAGCAACTCCTTAAGCGGCTCTATCAGCATCAT 3858
299 SerGlyGlySerLeuSerIleLeuAlaGlyLysSerValLeuPheGly 315
||| : | | | | | | | | | | | : | | | | |
3859 GCCTCAGAGAGCTCAGTCTTCTTGCAGACAGACAGAAACATTACCTTGT 3908
315 naSnAsnSerGlnArgThr...SerAspGlnGlyLeuValArgAsn 330
||| : | | | | | | | | | | | : | | | | |
3909 AAGAAATACCTTACACACACCGGAAATGACCTCTTAACGATATG 3958
330 laIleTyrLeuGlyLysAspAlaIleLeuSerSerLeuGlu...AlaArg 345
||| : | | | | | : | | | | | : | | | | |
3959 CGATCAACATGAGCAATACCGGAAATTCACCGCAATTCACGCTCCGAAA 4008
346 AsnGlyAspIleLeuPhePheAspProIleValGlnIleSerSerSer 362
||| : | | | | | | | | | | | : | | | | |
4009 AATCATACAAATTTCTTCTATGATCCCTACCTTACAGAGAACCTCATC 4058
362 sGluSerProLeuProSerSerIleuGlnAlaSerValThrSerProThr 379
||| : | | | | | : | | | | | : | | | | |
4059 AGAGTATTGAAGATAAATACGGCTCGCGAGCTCATCATCATCATATC 4108
379 roAlaThrAlaSerProLeuValIleGlnThrSerAlaAsnArgSerAl 395
||| : | | | | | : | | | | | : | | | | |
4109 AAGCAACG...ATT 4119
396 IlePheSerSerGluArgLeuSerGluGlu...LysThrProAspAs 411
||| : | | | | | | | | | | | : | | | | |
4120 CTATTTTCTGAGAAACCTTAACAGACAGATGACTTAAGTTGCGACAA 4169
411 nLeuThrSerGlnLeuGlnProIleGluLeuLysSerGlyArgLeuVal 428
||| : | | | | | : | | | | | : | | | | |
4170 TTTAAATATCTTCAATCAGCAGCGCTCTCCCTATCCGAGAGAAAGTTAT 4219
428 alLeuLysAspArgAlaValLeuSerAlaProSerLeuSerGlnAspPro 444
||| : | | | | | : | | | | | : | | | | |
4220 TCCTACAAAGAGAGTCACTTACAGACACACAGCTTCTCAACAGGCC 4269
445 GlnAlaLeuIleMetGluAlaGlyThrSerLeuLysThrSerSer... 460
||| : | | | | | | | | | | | : | | | | |
4270 GGTTCCTCTCGCATGATTCAGAGACGACATTAATCACTACAGCTGG 4319
461 AspLeuLysLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrG 477
||| : | | | | | : | | | | | : | | | | |
4320 GAGTTTACATACAGAACCTAGCAATCAATGATCTTACGCTTCA 4369
477 LuLysSerValThrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
||| : | | | | | : | | | | | : | | | | |
4370 AGCAGCCCTCAGCCTAACAGCAAAAGCTCTTCA...AATAAGCTGATC 4416
```

```
494 LeuSerAsn...SerGlyAspGlnAspPheArgLys 505
||| : | | | | | : | | | | | : | | | | |
4417 GTATCTGGGAAGCTCAACCTGATGATTTGACGGAACATTTATGAAAG 4466
505 nValGluLeuLeuSerLysGluGln... 513
||| : | | | | | : | | | | | : | | | | |
4467 T...CATATGCTTACAGCCATGACACAGCTCTTCTCTATTTAAATTCACGG 4513
514 .....AsnAsnIleProLeuLeuThrLeuSerLysGlu 524
||| : | | | | | : | | | | | : | | | | |
4514 TTGATGCTGATGTTGATATCAATGACCTGACATCAGCACCTT... 4554
525 GlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyr 541
||| : | | | | | : | | | | | : | | | | |
4555 .....ATCCCTGTTCTCGCTGAGAGATCTTAATCAGATATGCGATT 4595
541 rGlnGlyAspTyrPhePheSerThrLysAspSerAspGluLysHisSer 557
||| : | | | | | : | | | | | : | | | | |
4596 CCAAGACAAATGCAATGTTAAATGACTACGAGAACAGCTACAAATACAA 4645
558 .....LeuIleAlaAsnThrPheProLysAsnTyrValProHisProGlu 572
||| : | | | | | : | | | | | : | | | | |
4646 AAGAGCCACCGCAACTTGCACCAAAACAGAGATTGTTCCACGCCCGAA 4695
573 ArgGlnSerThrLeuValAlaAsnThrLeuThrPheAsnThrTyrSerAspMe 589
||| : | | | | | : | | | | | : | | | | |
4696 AAGAAATCTGCGTTAGTATGCAATACCTATGGGAGCTTTACTGACAT 4745
589 rGlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrL 606
||| : | | | | | : | | | | | : | | | | |
4746 TCCTCTCTGCAACAGCTTGTGAGAGATCGGCCCAACTGTTGACAAACA 4795
606 ePheGlyThrTyrGlySerAlaValSerAsnLeuPheTyrAlaHisAsp 622
||| : | | | | | : | | | | | : | | | | |
4796 AACAGGTTCTGCTTCCATGACGAAAC...TTCCCTCAATAG 4839
623 SerSerGlyLysProIleAspAsnThrHisArgSerLeuGlyTyrLe 639
||| : | | | | | : | | | | | : | | | | |
4840 ACTGAGAGTGAATTCGCAAGGCTTCGCTATACCTTGGAGGCTACGT 4889
639 uPheGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAla 656
||| : | | | | | : | | | | | : | | | | |
4890 CATCGTGGAAGTCTCAGACTCTTAAGACGACCTATTACCTTTCGCT 4939
656 laGlyGlnLeuLeuGlyLysSerSerAspPheIleThrSerThrGlu 672
||| : | | | | | : | | | | | : | | | | |
4940 TCTGCCATCTTGTGCTAGAGACAAAGATGTTTATCGCTCAGACAAAC 4989
673 ThrThrSerTyrIleAlaThrVal... 680
||| : | | | | | : | | | | | : | | | | |
4990 TCTAGAACCTACGGTGGAACTTATTCTTCAGACACTCTCATACCCTACA 5039
681 .....GlnAlaGlnLeuAlaThrSerLeuM 689
||| : | | | | | : | | | | | : | | | | |
5040 ACCCCAAACTATTGAGATTAGAGAGCAAGTTTCTTAACAGCTA 5089
689 eLys...IleSerAlaGlnAlaCys 696
||| : | | | | | : | | | | | : | | | | |
5090 TAGAAAAATTCCTAGGAAATTCCTCAGCTTGAGATGTCAGAGTTTCG 5139
697 TyrAsnGlnSerIleHisGluLeuLysThrLysTyrArgSerPheSer 713
||| : | | | | | : | | | | | : | | | | |
5140 TTCAGCCATTGAGAACCGTATGAAAGCAGCATATACCTGATTCG...CC 5186
713 sGluGlyPheGlySerThrPheHisSerValAlaValSerGlyValCys 730
||| : | | | | | : | | | | | : | | | | |
5187 AGAATCCCAAGCTTCTGACCAACGAGTGTATAGCTGTGCTATCCGCC 5236
730 laSerIlePro...IleValSerAsnGlySerGlyLeuPheSerSerPhe 745
||| : | | | | | : | | | | | : | | | | |
5237 TAGACCTTCTTCTTGTCTTTCACCAACCCACATCTCTTTCACAGCCTTC 5286
```

```

1670 AGCAGCCCTCAGCCTAACAGCAAAAGCTCTTCA...AATAAGTCATC 1716
494 LeuSerAsn.....SerGlyAspGluAsnPhetrgLias 505
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1717 GTATCTGGAGACTCAACCTGATTTGATTTGAAGGAACTTTATGAAG 1766
505 nvaGluLeuLeuSerIysGluGln..... 513
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1767 T...CATATGTCAGCCATGACCAAGCTCTTCTCTATTTAAAAATCAGG 1813
514 .....AsnAsnIleProLeuLeuThrLeuSerIysGlu 524
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1814 TTGATGCTCATGTTGATACCTTACCTTACATTCAGCACCTT..... 1854
525 GlnSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTy 541
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1855 .....ATCCCTGTTCTGCTGTAGAGATCCCTAATTCAGAAATACGGAT 1895
541 tGlnGlyAspTrpThrPheSerTrpLysAspSerAspGluGlyHisSer 557
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1896 CCAAGGCAATGGAAATGTTAATTTGGACTACGGATACAGTACAAATACAA 1945
558 .....LeuIleAlaAsnTrpThrProLysAsnTyValProHisProGlu 572
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1946 AAGAGGCCAGCGCAACTTGGACCAAAACAGAGATTTGTTCCACGCCCGCA 1995
573 ArgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTrpSerAspMe 589
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1996 AGAAATCTCGCTAGTATGCAATACCTTATGCGGAGCTTTACTGACAT 2045
589 tGlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyValAlaTrpL 606
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2046 TCGCTCTCTGCACAGCTTGAGAGATCGGCGCACTGGTATGGAAACCA 2095
606 eupheGlyThrTrpGlySerAlaValSerAsnLeuPheTyAlaHisAsp 622
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2096 AACAGAGTTTCTGGTTCCTCCATGACAGAAC.....TTCTCCATATAC 2139
623 SerSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyLe 639
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2140 ACTGAGATCGAAATCGCAAGGCTTCGCTATCCTCGAGAGGCTACTG 2189
639 upheGlyLysSerThrHisSerLeuAspAsnHisSerPheCysLeuAla 656
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2190 CATGGGTGGAAAGTGTCACTCCTTAAGAGCAGCTATTTCCTTTGGCT 2239
656 IagGlnLeuLeuGlyLysSerSerAspPheIleThrSerThroL 672
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2240 TCTCCATCTCTTTGCTAGACAAAGATGTTTATCGCTCACACAAAC 2289
673 ThrThrSerTyrlleAlaThrVal..... 680
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2290 TCTAGAACCTACGGTGGAACTTTATCTTCAAGCACTCTCATACCTACA 2339
681 .....GlnAlaGlnLeuAlaThrSerLeuM 689
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2340 ACCCAAACTATTGAGATTAGAGAGCAAGTTTCTGCAATCAGCTA 2389
689 etLys.....IleSerAlaGlnAlaCys 696
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2390 TAGAAAAATTCCTAGAGCAATTCGCCCTAGCCTTCGATGCCAAGTTCC 2439
697 TyrAsnGlnSerIleHisGluLeuLysThrLysTyArgSerPheSerLe 713
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2440 TTCAGCCCTTACAGACACCTATGGAAAGCAGCTATACCTCATTT...CC 2486
713 sGlnGlyPheGlySerTrpHisSerValAlaValSerGlyValuValCys 730
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2487 AGAATCCAGAGTTCTTGGAGCAAGAGTGTATACCTGTTATCGGCC 2536
730 IasertIlePro...IleValSerAsnGlySerGlyLeuPheSerSerPhe 745
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2537 TAGACCTTCCTTTGTTCTTTCACACCAACATCTCTTTTCAAGACCTTC 2586

```

```

746 SerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheG 762
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2587 ATTCCACAGATGAAAGTCGAATGCTTTATGTTACAAATATGCTTCT 2636
762 uGlnSerSerGlyGluIleArgSerPheSerAlaSerSerPheArgAsn 779
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2637 CGAAGCTCTAGATGAGCCGCTGTTTACTATTGGAAAGCTCTGTTAAC 2686
779 IeserLeuProIleGlyIleThrPheGluLysLysSerGlnLysThrArg 795
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2687 TCTCGAATTCCTTCGCGCCAAATTCCTGCACGCGCATATCGAGATTC 2736
796 ThrTrpTyTrpPheLeuGlyAlaTrpIleGlnAspLeuLysArgAspVa 812
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2737 TACACCTATGATCTTCAGGATTTCTGTTCCGATGCTATGATACAA 2786
812 IglSerGlyProValIleLeuLeuLysAsnAlaValSerTrpAspAla 829
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2787 TCCCCATTCACACGCACTCTTCGATGACCCGCACTCTTGAAATTC 2836
829 roMetAlaAsnLeuAspSerArgAlaTrpMetPheArgLeuThrAsnGln 845
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2837 CCGGTGCGCAATCTTTCAGACAGCATTTTACTACGAGGCTAGCCAAC... 2883
846 ArgAlaLeuHisArgLeuGlnThrLeuLeuAsnValSerCys..... 859
2884 .....ACTACGCTACACATCCCAATTTGAGCTTT 2915
860 .....ValIeuArgGlyGlnSerHisSerTySerLeu 871
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2916 CGGACATTACGCTATGCAACCTCGCTGCATCTTCAGACAGTATGTAC 2965
871 sPLeuGlyThrThrTrpArgPhe 878
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2966 ATGTGTGATCCCAACTCGATTC 2988
seq_name: gb_ba2:CPOMP54
seq_documentation_block: 6030 bp DNA BCT 22-JAN-1999
LOCUS CPOMP54
DEFINITION Chlamydia pneumoniae omp5 and omp4 genes.
ACCESSION AJ001311
VERSION AJ001311.1 GI:3255934
KEYWORDS omp4 gene; omp5 gene; outer membrane protein 4; outer membrane protein 5.
SOURCE Chlamydia pneumoniae.
ORGANISM Chlamydia pneumoniae
REFERENCE 1 (bases 1 to 6030)
AUTHORS Knudsen, K.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-1997) Knudsen K., Department of Medical Microbiology and Immunology, University of Aarhus, The Bartholin building, University of Aarhus, DK-8000 Aarhus C, DENMARK
2 (bases 1 to 6030)
AUTHORS Knudsen, K., Madsen, A.S., Mygind, P., Christiansen, G. and BirkeLund, S.
TITLE Identification of two novel genes encoding 97- to 99-kilodalton outer membrane proteins of Chlamydia pneumoniae
JOURNAL Infect. Immun. 67 (1), 375-383 (1999)
MEDLINE 99081766
FEATURES
     source
         1..6030
         /organism="Chlamydia pneumoniae"
         /isolate="CDC/CWL-029/VR-1310"
         /db_xref="taxon:83558"
         /clone="PEX2-18"
         /clone="B8F3"
         /clone_1lb="pbluescript"
         /clone_1lb="PEX2"
         /dev_stage="elementary body"
         complement(1..2804)
gene

```

22128 AAGATGAGTGTACTTACTTCCAAATACTATACCGAGTCGAGGATCG 22177
 447 LeuLeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAsp...Le 462
 22178 AAAGTCGTATGATGAGGAGGACTCTTTGAGCGACGCGCTGAGGGGT 22227
 462 uLysLeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGlu 478
 22228 CAGCTCAATGGCTAGCCATTAATAGATTCCTAGATGGGAAATA 22277
 478 ySerVal...ThrIleHisAlaProAsnLeuSerIleGlnLys 491
 22278 AAGCATCATTAAGCGCGACGACCAACATGACATGTCCTTATACGAGG 22327
 492 IlePheLeuSerAsnSerGlyAspGluAsnPheTyrgluAsnValGlu 508
 22328 CCTATCATGCTTGTAGATGCTCAGGGGAACTATATAGCATCATAACT 22377
 508 uLeuSerLysGluGlnAsnAsnIleProLeuLeuThrLeuSerLysGlu 524
 22378 C...AGTCACAGCAGGCTTCTCTTAATAGAGCTTCTGACACAG 22421
 525 GlnSerHisLeuHisLeuProAspGly...AsnLeuSer 536
 22422 GAAGGATGACACTAGACATATCCCGCATACCCCAATTCCTAAATACTACG 22471
 537 SerHisPheGlyTyrglnGlyAspTrpThr...PheSerTrpLysSerAsp 553
 22472 AATCACTATGGGTATCAAGGAGACTGMAATATGTTGGTGCAGCATG 22521
 553 spGluGlyHisSerLeuIleAla...AsnTrpThrProLysAsnTrp 567
 22522 CAAGTCGAAAAAATAAAATGCTACCTTAAGTGGACATAAAGAGATG 22571
 568 ValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAs 584
 22572 AAGCGAATCCAGAACGTCGACGACCTTGGTCTTAATACCGTGTGGG 22621
 584 nThrTySerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 601
 22622 TTTCTTGTGATGTCGCGCTCCATTCAGAGCCATGAGCGGAGCACAA 22671
 601 IsGlyGlyAlaTyrlLeuPheGlyThrTrpGlySerAlaValSerAsnLeu 617
 22672 GTTGTGATATCTCTCAACAAATTTGGGTATCAGGAATGCGGAC... 22718
 618 PheTyAlaHisAspSerSerGlyLysProIleAsnTrpHisHisAsr 634
 22719 TTTTTCATGAGATCAGAAAGCAACCA...CGTAGTTATCGCATTC 22765
 634 gSerLeuGlyTyrlLeuPheGlyIleSerThrHisSerLeuAspAspLys 651
 22766 TAGCGCGGTATGATAGAGAGAGATTCCTCAGCGCTCTGAAAT 22815
 651 erPheGlyLeuAlaAlaGlyGlnLeuGlyLysSerSerAspSerPhe 667
 22816 TCTTAAATTTGCTTTTGTGCTTTTGGCTACGAGCAAGACCACTT 22865
 668 IleThrSerThrGluThrThrSerTyrlIleAlaThrValGlnAlaGln 683
 22866 CTGGCTAAGAACCATACCATGTTATTCAGAGCGCAATGACTTACCCACA 22915
 684 LeuAlaThrSerLeuMetLysIle... 691
 22916 CTTGGAGAGCTTAAGACCCCTGCTAGATTTTGTCAAGAAATTCGACT 22965
 692 SerAlaGlnAlaCysTyrglnGluSerIleHis 702
 22966 CCTACCTTTTGTCTTCAATGCTGTGGTTTGGCTTATGGCCATACGACAACT 23015
 703 GluLeuLysThrLysTyrglnSerPheSerLysGluGlyPheGlySerTr 719
 23016 AACATGACCAACAAAGTACACGCTATTCCTCCTGTAAG...GGAACCTG 23062

719 phISerValAlaValSerGlyGluValCysAlaSerIleProIleValS 736
 23063 GGAATATGATGCTTCGGTATAGATATGAGAGAGACTTCCCGGAGTGG 23112
 736 erAsnGly...SerGlyLeuPheSerSerPheSerIlePheSerLysLeu 751
 23113 CTTACGAGAGCTGCGCTCTGGTGCATACCCACACCGCATTTCTTAACCTA 23162
 752 GlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerGlyGluI 768
 23163 GAGATGATCTATAGCACATCAGAAATGACTTTAAGGAAACGCGACAGAGG 23212
 768 earGSerPheSerAlaSerSerPheArgAsnIleSerLeuProIleGlyI 785
 23213 CGGTCTCTTCCAAAGTGAAGACCTTCTCAATCTACCGGTTCTCTGAGGA 23262
 785 leThrPheGluLysLysSerGlnLysThrArgThrTyrglyTyrlPheLeu 801
 23263 TAAAAATTTGAGAAATTCCTCCATTAAGTCT...AGGTATGATCTCTCAWA 23309
 802 GlyAlaTyrlIleGlnAspLeuLysArgAspValGlnSerGlyProVal 818
 23310 ...GCTTACGTTCCCGATGATGATTCCTAATGATCCAGGCTCCACGACAC 23356
 818 IleuLeuLysAsnAlaValSerTrpAspAlaProMetAlaAsnLeuAspS 835
 23357 TCTATGATGTTCTGGGAGATTCCTGGTGCACATGAGTCAAGCTTGCTA 23406
 835 erArgAlaTyrlMetPheArgLeuThrAsnGlnArgAlaLeu...HisArg 850
 23407 GACAAAGCTCTCTGTACGTCCTGGAATCATCATGCTTTCCTTCAAC 23456
 851 LeuGlnThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHis 867
 23457 TTTGACGTTTTCAGTCACTGTTGAAGTCAAGTTCGAGGATCTTCTCGTAG 23506
 867 rTyrlSerLeuAspLeuGlyThrThrTyrglyPhe 878
 23507 CTATGCTATGATCTCTGGAGAGATTCGATTT 23540

seq.name: gb.pat1:A81827

seq.documentation_block:

LOCUS	A81827	DNA	PAT	21-JAN-2000
DEFINITION	Sequence 1 from Patent WO9858953.			
ACCESSION	A81827			
VERSION	A81827.1 GI:6731863			
KEYWORDS	unidentified.			
SOURCE	unidentified.			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 3200)			
AUTHORS	Madsen A. and Birke Lund, S.			
TITLE	NOVEL SURFACE EXPOSED PROTEINS FROM CHLAMYDIA PNEUMONIAE			
JOURNAL	Patent: WO 9858953-A 1 30-DEC-1998;			
	MADSEN ANNA SOFIE (DK); BIRKE LUND SVEND (DK)			
FEATURES	Location/Qualifiers			
source	1..3200			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
CDS	205..>2987			
	/note="unamed protein product"			
	/codon_start=1			
	/protein_id="CAB69199.1"			
	/db_xref="GI:6731864"			
	/translation="MKTSLPWLVSSVLAFSCHIOSLANELLSPDSEFNIGIDSCF			
	TPKTSATVYSLNGDVEFPEPKGTPLSQCFQOTDNLFLNGHSLTFPCIDAGCTHA			
	GAASTIANKNLTFSGFSLSDSPSTVTVTGGTGLSSAGVLENIKRLVAVGNFS			
	TADGAIKASFLTGTSGLDALESNNSSTKGAIAATAGARIANNIGVRLSNIAS			
	TSGAIDDEGTSLLSNKRLYPEGNAARTGGAINTAASGPETLISNNKRLIFASN			
	VAEPSCGAHAKKALSSGGTFEPLRNWVSATPVGCAISIDASELSISATGNITF			
	VANTLFTTGSTDTFPRNAINTINISNGKFTELRAKHHITTFYDPIITSEGTSSDVLIKNN			

[illegible]

[illegible]

VSSLOLATERKRVSESLPELSLITKEHMYIAKLJENGAKKRTKRLSTSKSNAHC
 POKLILLENPAERTIHGRLPKRELPSMPTAYVLAMPNPDTSKXHGRLDKMLBLS
 SGLSEISELVANMKODEDLSTLILIRCKGKKEKRLIPTYNSIOWIQLYLNPKRLRE
 KDPFAITEENREGKRISTRSDRSQFKEKRLHSGHPTPIRHTIATWHELSGMDLK
 TIOALLGSSLETTIVTVOVSKRLOKHOEAHPHA
 4710 a 330 c 3665 g 4743 t
 BASE COUNT
 ORIGIN

alignment_scores:

Quality:	852.00	Length:	812
Ratio:	1.635	Gaps:	30
Percent Similarity:	64.163	Percent Identity:	31.158

```
alignment_block:
```

US-09-677-752-4 X AE001587

Align seg 1/1 to: AE001587 from: 1 to: 16448

[illegible]

994 ..CTTGTATTCTGAAATAGTGCACGACCGCAGCAAAATGAGCA 1041
232 AAlleCysCysIleSerAspLeuSnrThrSerGluLysGlySerLeuSe 248
1042 GCTCTTTCTGGAGATGCCGATGTC 1067
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
1068 CATATCTGGAAATCAGAGTGTACTTCTCAGGAAACCAAGCTTAGCTTA 1117
265 yGlyGlyAlaIleLeuPheAlaLysHsMeValLeuAlaGlyTyr.....Asn 279
1118 ATGGCGGACCCATTTATGTAGAACCTTACACTGCGCTCCGGGGGGGG 1167
280 GlyProValSerPheIleAsnAsn.....SerAlaLysIle 291
1168 GGGGATATCTCTTCTTCAACAATATAGTCAAGGATACCACTGACAGTAA 1217
291 eGlyGlyAlaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAlaG 308
1218 TGTGTAGCCATTTCTATCTAGTCGACAGTGTAGTACTTTTCAGCAG 1267
308 LylGluLysSerValLeuPheGlnAsnAsnSer...GlnArgThrSerAsp 323
1268 AAGCAGGACATCTACTCATTCGCAATGCCATTCCTTCTGACACTACACA 1317
324 GlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSe 340
1318 CAACACTACAAAGAAATCTATGTACATAGATCTACTGCAAAAGATCAC 1367
340 rSerLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProIleV 356
1368 GAATTTACTGCAATATCTGGGCATAGCATCTTTCTTACGATCCGATTA 1417
356 aGlnGlnLysSerSerLysGlnSerProLeuProSerSerLeuGlnAla 372
1418 CTCCTAATACGGCTCGCGATCTTACAGATCTTAAATCTCAATAAGCTT 1467
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThr 389
1468 GATGACGATAAT.....AGTAC 1484
389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGluG 406
1485 AGATTTATAGTGGTCGATGTTTCTTGTGTAAGCAAGCTCTCTGAAAGAG 1534
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnProIleGlu 421
1535 AAGCAAAAGCTTCAGACACACTCTACGCTGACAGCAGCTCTACT 1584
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPr 438
1585 CTAACTGCAGAAATTTACTACTTAACGTGTGTCACCTCGATACGAA 1634
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyHrs 455
1635 AGGCTTTACTAGACCGCGGTTCTCTGTTATTAATGATGCGGACACA 1684
455 eLeuLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIlePro 470
1685 CATTAAACCAAGTACAGAGAGCTACTTAAACAGCTCTTCCATTCTCT 1734
471 LeuHsSerLeuAspThrGluLysSerValThrIleHsAlaPro..... 485
1735 GTAGACTCTTAGCGGAGGTAGAAAGTTGAATTTGCTTCGACAGC 1784
486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
1785 AAGTAAAAATGTAGCCCTTAGTGTCGATTTCTTTTGGATAACCAAG 1834
498 lYAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlnGlnAsn 514
1835 GC.....AATGCTTATGAATAATCACGAC...TTACGAAAAACCTCAAGAC 1875

515AsnIleProLeuLeuThrLeuSerLysGlnGlnSerH 527
1876 TTTCATTTGTGCAGCTCTGCTGCTGGTACTGCACAACTACAGATGT 1925
527 sLeuHsLeuProAspGlyAsnLeuSerSerHsPheGlyTyrGlnGlyA 544
1926 TCCACCGCTTCTACAGTACGAACTCTACGACATATGCTATCAAGGTA 1975
544 sPThrPheSerThrLysAsp.....SerAspGluGlyHs 556
1976 CTGGCGAATGACTTGGTTCATGATACCCGACAGCTCCAAAGACTAAG 2025
557 SerLeuIleAlaAsnThrPheProLysAsnTyrValProHsProGluAr 573
2026 ACAGGACATTAGCTTGGACCAATAGAGCTACCTCGGAATCTCGACG 2075
573 gGlnSerThrLeuValAlaAsnThrLeuThrAsnThrTyrSerAspMetG 590
2076 TCAAGACCTTATAGTCTCAATACCTTTGGGAGATCTTTTCAGACATCC 2125
590 lAlaValGlnSerMetIleAsnThrIleAlaHsGlyAlaTyrLeu 606
2126 AAGCATTCAAGCTGTCTATAGACAGACTGCTTGTGACTTTGTTACAT 2175
607 PheGlyThrTyrGlySerAlaValSerAsnLeuPheTyrAlaHsAspSe 623
2176 CGAGCTTCTGGCTGCGCGAGTCCCAAT...TTCTTAGTAAGAATAA 2222
623 rSerGlyLysProIleAspAsnThrPheHsIleArgSerLeuGlyTyrLeuP 640
2223 GAAAGGGAATAA...CCCAATATCCGTCAATAATCTGTGGAATATGCTA 2269
640 heGlyLysSerThrHsSerLeuAspAspHsSerPheCysLeuAlaAla 656
2270 TCGGAGGTCCAGCGCAACTGTTGTGAACCTTAATTAAGCTTTGCTTT 2319
657 GlyLysLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
2320 TCCCAACTCTTGTAGAGGATAAAGATTCTTAGCGCTAAATAATCATAC 2369
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
2370 TGATATCATAGCAGAGACTTCTATATCCAAACATTTACAGAAATGATGTG 2419
688 688
2420 GGTTCATAGCTTCTCTTACATAACTCTCGCTTGGAGTCATATAA 2469
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGlnSerIleHsGluLe 704
2470 CCCCTCGTTTAGAAGGACAGCTCGCTTATAGCCAGCTACATATGATCT 2519
704 ulysThrLysTyrArgSerPheSerLysGluGlyPheGlySerThrHs 721
2520 GAAGCAAAAGTATACCTCGTAT...CTGAGAGTGAAGAGCTTTGGGGCA 2566
721 eValAlaValSerGlyLysValCysAlaSerIleProIleValSerAsn 737
2567 ATATTCCTTTTAACTGATGTTGGAGCTTCTTCTCATTTTATCTTAA 2616
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
2617 TACGTGATGTTTGTATACCTATAGCTCATACATCAAGATGTGAC 2666
754 eSerGlyThrGlnAspGlyPheGluGlnSerSerGlyGluLeuArgSerP 771
2667 CTATATACGTAGAGACACTTCTCGAAGAAAGTACAGAACAGAGTCTT 2716
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
2717 TTGATGACGACACACTTTCATATTATCTTTCCTATATAGGGGTAACTTT 2766

```
633 ACGGGAAGTCTTTTTCGACAACTTCGATTCGACCTGCAA 682
206 IecysCysSerAsnLeuIleCysSerGlyAsnValAsnProLeuPhePhe 222
    |||
    |||
683 TTTTACAGAAAGCTCGTGACTTTCTTAATAAGTAAAGTTCTT 732
223 ThrCysAsnSerAlaThr.....As 229
    |||
    |||
733 ATTGACAATAAGTACACAGAGAGCTCTCAACAACGGGGATATGTC 782
229 nglyGlyAlaIleCysCysIleSerAspLeuAsnThrSerGlyLysGly 246
    |||
    |||
783 AGAGCTGCTATCTGTGCT.....TATAAACTAGTACAGATCTA 823
246 erLeuSerLeuAlaCysAsnGlnGlnThrPheAlaSerAsnSerAla 262
    |||
    |||
824 AGGTCACTCCTCAGTCAAGATCAGATGTTACTCTTCACACAAATACATCG 873
263 LysGlyLysGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAs 279
    |||
    |||
874 ACACACCGGAGAGAGCTATCTATGTGAAAAGCTGACCTGGCTCGG 923
279 nglyProValSerPheIleAsnAsnSer.....AlaLysI 291
    |||
    |||
924 AGCACTTACCTTACGATGAGAAATAGTCAATGAGAGTACACCTCTA 973
291 IeglyGlyAlaIleAlaIleGlnSerGlyLysSerLeuSerIleLeuAla 307
    |||
    |||
974 AAGTGAAGCCATAGCTATCGAAGATAGTGGGAGATTGAGTTTCCGCC 1023
308 GlyLysLysSerValLeuPheGlnAsnAsnSerGlnArgThrSerAspG 324
    |||
    |||
1024 GATAGTGTCACATGTTCTTTTAGGAAATACAGTCACTCTACTCTCC 1073
324 nglyLeuValArgAsnAlaIleTyrLeuGlyLysAspAlaIleLeuSer 341
    |||
    |||
1074 TCGGACCAATAGAGATGATTCACCTTACGAGACCAAGTCACAG 1123
341 erLeuGlyAlaArgAsnGly...AspIleLeuPheAspProIleVal 356
    |||
    |||
1124 CTTCGCTTCTGCTGCTGTAGAGCCATCTACTTCTATGATCCCATACT 1173
357 GlnCysSerSer.....LysGlnSerPr 365
    |||
    |||
1174 ACAGATCTTCCACACAGTACAGATGCTTAAAGTAAATGAGACTCC 1223
365 OleuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrA 382
    |||
    |||
1224 GCCAGATTCGCACTACATATACA..... 1248
382 IaserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 398
    |||
    |||
1249 .....GUGAACATCATCTTCA 1266
399 SerCysLeuArgLeuSerGlnGlnGlyLysThrProAsp...AsnLeuThrSe 414
    |||
    |||
1267 GGAGAAAGTATATCAGAGACAGAGCCGAGATTCCTAAATCTTACTTC 1316
414 rGlnLeuGlnGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysA 431
    |||
    |||
1317 GAAGCTACTACAGCTGTACTCTTTCACAGAGTACTATCTTAAAC 1366
431 sPArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
    |||
    |||
1367 ATGAGAGTACTGTCAGACTCAGGACATTCACCTACACAGCAGATTCGCT 1416
448 LeuIleMetCysAlaGlyThrSerLeuLysThrSerSerAspLeuLysIe 464
    |||
    |||
1417 CTCGAATGACGTAGAACTACTCTAGAACCTGATCTACTAGCACCAT 1466
464 uAlaThrLeuSerIleProLeuHisSerLeuAspThrGlnLysSerValT 481
    |||
    |||
1467 AACCAATTTGCTATTAACTCACTGATCTATAGACCGCTGCAAAAGACAA 1516
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
    |||
    |||
1517 AATAGAACCAAAAGCTACTGCTCAAAAATCTGACTTATATGGAACCATC 1566
494 LeuSerAsnSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSe 510
    |||
    |||
1567 ACTTATTATGACCCGAGCGGACAGCTTTTATGAAATCATGTTTAAGAAA 1616
510 rLysGlnGlnAsnAsnIleProLeuLeuThrLeuSer...LysGlnGlnS 526
    |||
    |||
1617 TCCCACTCTTACAGACATCTTACAGCTCAAAAGCTTCGAACTGTAACAA 1666
526 erHisLeuHisLeuProAspGlyAsnLeuSer.....HisPheGly 540
    |||
    |||
1667 GCACCGCAGTACTCCAGATCTATATGAGTGAAGAAATTCATTACGGC 1716
541 TyrClnGlyAspThr...ThrPheSerThrLysAspSerAspGlyLysI 556
    |||
    |||
1717 TATCAGGCACTTCGGGCCCAATGTTTGC.....GGCACAGCGGC 1757
556 sSerLeuIleAla.....AsnThrPheProLysAsnTyrValProHisP 571
    |||
    |||
1758 TTTCACACCTGCAACCTTACACGTGACATAAAGCTGATATATCTAATC 1807
571 rGlnArgGlnSerThrLeuValAlaAspThrLeuThrPheAsnThrTyrSer 587
    |||
    |||
1808 CCGAGCCTATCGCCTCTTATGCTCCTATATAGCTTATGAGATGATTTATA 1857
588 AspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
    |||
    |||
1858 GATATTATGCTCTCCATTCATTCATGACAGACGCAACAGACGTTGCA 1907
603 GlyAlaTyrLeuPheGlyThrTyrGlySerAlaValSerAsnLeuPheT 619
    |||
    |||
1908 GCGAGACCTGCTTTT.....TGCTCTCGTGGATATATCAACTCTTC. 1950
619 rAlaHisAspSerSerGlyLysProIleAspAsnThrPheHisIleAspSer 635
    |||
    |||
1951 ....CATAGATAGTACAAAACACGACGCGGTTTCCGCCATTTGAGT 1995
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspHisSerPh 652
    |||
    |||
1996 GCGCGTATATGTCATACAGACGAAACCTACATCTTTCACATTAAGATTC 2045
652 eCysLeuAlaIleGlnLeuLeuGlyLysSerSerAspSerPheIleT 669
    |||
    |||
2046 TATGCTGCAATTTGTCAGCTCTTTCGAGAGATAGACACTTGTGTAG 2095
669 hrSerThrGlnThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAla 685
    |||
    |||
2096 CTAGAGTCAAGGTACAGTCTACGAGAGAACTCTATTTACACGACACAC 2145
686 ThrSerLeuMetLysIle..... 691
    |||
    |||
2146 GAACCAATATCTCTCTCTGCTCAAACTACGCGCTTGTGTTGTTTAA 2195
692 .....SerAlaGlnAlaCysTyrAsnG 699
    |||
    |||
2196 TGTTCCTACAGAGATTCCTGTTCTCTTTCAGGAAACCTTACGACACC 2245
699 IuSerIleHisGlnLeuLysThrLysTyrArgSerPheSerLysGlnGly 715
    |||
    |||
2246 ATACGGATACGATCTGAAACCAAGTATACACATATCTTACTGTTAA 2295
716 PheCysThrPheHisSerValAlaValSerGlyGluValCysAlaSerI 732
    |||
    |||
2296 ...GGAAGCTGGGGAATGATGATTCGCTTTCGAAATTCGGTGAAGAGC 2342
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheS 749
    |||
    |||
2343 TCCGATTCCTTAGATGAAGATGCTCTATTTGAGCAGTACATCCCTTCA 2392
```

```
12 CysAlaIleLeuSerThrAlaValLeuPheGlyGlnAspProLe 28
|||||
7593 TGTCGACTGTTTTCGCAAGCTGCTGCAAAATATAGGCCCTCTGATAG 7544
28 u.....GlycylThrAlaLeuLeuThrLysAsnProAsnHisValC 43
|||||
7543 CTTCGCGAGAGTACTACACGACGACCTATCTCTAAATAATACGATA 7494
43 ys.....ThrPheGluAspCysThrMetClnSerLeu... 54
|||||
7493 CTGCAATGACTATCTCTGACAGAGATATATGCTGCAAAACCTTGGG 7444
55 ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
|||||
7443 GATTCGCGAGCTTTTAACGAGCGCTGCTTTTCTGACACTACGAGATCTTT 7394
68 uTrpValLeuLeuLysSerThrCysTrpPheValSerLysLeuHisL 85
|||||
7393 AAGCTTTCGCGTAGGGGTACTCA.....CTTCTTTTAAATATATA 7350
85 hraspProLysGluAlaLeuPheLysGluLysCysAspLeuSerLecGln 101
|||||
7349 AGCTAGTCTGTAAGGC.....GACGACCTTCTCTGTATACA 7315
102 AsnPheArgPheLeuSerPheThrAspCysSerLysGlu.....Se 116
|||||
7314 ACTGATAAATATCTGCTCCCTAACAGCATTTTCCACTCTTCTCTACG 7265
116 rSerProSerLecLeu.....HisGlnL 124
|||||
7264 GGCCCATCATCGGTATACACACCCTCTCGAAGAAAGGTGACGTTAAT 7215
124 ysAsnGlyGlnLeuSerLeuArgAsnGlySerMetSerPheCysArg 140
|||||
7214 CTGGAGCGGATCTTACATTTGATTAACAATGACATATTTTATTAACAA 7165
141 AsnHisAlaGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLe 157
|||||
7164 CATTACTGTGAGAAATCCGCGAGCCATTTCTACCAACAATCTTCTCT 7115
157 uGlnHisAsnTrpLeuPheThrAlaPheGluLysSerSerLysGly 173
|||||
7114 GAAAAACGACGCGGATGATTTCTTTGAAAGGAAATTAATCGACGCCAA 7065
174 .....AsnGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
|||||
7064 CAGGAAATAAGGTGGGCTATTTGCTCTACTGCTACTGATATATATACA 7015
187 ArgAsnValSerProLysPheAlaArgAsnArgAlaAspLeuAsnG 203
|||||
7014 AATATACGCGCTCTACCTCTCTCTCGAACATATTCCTGACGTCACG 6965
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
|||||
6964 TGGACCTTAAATAGCACAGAACTGTACATTAACGGAATACGCTCT 6916
219 rLeuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
|||||
6915 ..CTTGTTTCTGAAATAGGTGACACGACCGCAGAAATGACAGA 6868
232 AlaIleCysCysIleSerAspLeuAsnThrSerGlyLysGlySerLe 248
|||||
6867 CCWCTT.....TCTGACATCCCGAGATGTAC 6842
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysG 265
|||||
6841 CATATCTGGAGTACAGAGTGTACTTCTCAGAAACCAACGCTGAGCTA 6792
265 ySGlyAlaIleLeuThrAlaLysHisMetValLeuArgTyr.....Asn 279
|||||
6791 ATGGCGAGCACTTATGCTAAGAAAGCTTACACTGCG..TTCCGGGGGG 6743
280 GlyProValSerPheIleAsnAsn.....SerAlaLysIle 291
|||||
6742 GGGGATATCTCTTTTCAACAATATGCTCCAGGTACACTGCAAGGTA 6693
291 eGlyGlyAlaIleAlaIleGlnSerGlyCysLeuSerLecLeuAlaG 308
|||||
6692 TGGTGAGACCATTTCTATATGCGACGTGAGAGTGTAGTCTTTCAGAG 6643
308 LysGlyLysValLeuPheGlnAsnAsnSer...GlnArgThrSerAsp 323
|||||
6642 AAGCAGCGGACATTAACCTTCAATGGGAATGCCATTTGTGCAACTACA 6593
324 GlnGlyLeuValArgAsnAlaIleTrpLeuGluLysAspAlaIleLe 340
|||||
6592 CAACATACAAAAAGAAATCTATTGACATAGATCTACTGCAAGATATAC 6543
340 rSerLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProLe 356
|||||
6542 GAATTTACGTCAATATCTGGGCAATAGCATCTTTTCTACGATCCGATTA 6493
356 aGlnGlySerSerLysGlnSerProLeuProSerSerLeuGlnAla 372
|||||
6492 CTGCTAATACGCGTGGGATTTCTACAGATCTTAAATCTCAATAGGCT 6443
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleG 389
|||||
6442 GATGAGGATAT.....AGTAC 6426
389 rSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluG 406
|||||
6425 AGATTATAGTGGTGATGTTTCTTTCGTGTAAGAGCTCTGTAAGATG 6376
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIle 421
|||||
6375 AAGCAAAAGTTGACAGACACCTCTTACCGTGAACACACCTGTACT 6326
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAla 438
|||||
6325 CTACTCGAGGAATTTGACTACTTAACGTGTGTCTGCTCGATACGAA 6276
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetClnAlaGly 455
|||||
6275 AGGCTTACTCGACACCGCGGCTCTCTGTTATATGATCGGCGACAA 6226
455 eLeuLysThrSerSer...AspLeuLysLeuAlaIleThrLeuSerIle 470
|||||
6225 CTTTAAAGCAGATACAGAGGCTCACTTAACAGCTCTTCCATTCCT 6176
471 LeuHisSerLeuAspThrGlyLysSerValThrIleHisAlaPro... 485
|||||
6175 GTAGACTTTTAGCGGAGGTAAGAAGTGTATGTGCTGCTTCGACG 6126
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
|||||
6125 AAGTAAATATGACCCCTTACTGTCGATTCCTCTTTTGGATTAACCA 6076
498 LysAspGluAsnPheTrpGluAsnValGluLeuLeuSerLysGluGln 514
|||||
6075 GG.....AATGCTTATGAAATATCACGAC...TTAGAAATACTCAAGAC 6035
515 .....AsnIleProLeuLeuThrLeuSerLysGlnGlnSerH 527
|||||
6034 TTTTCATTTTGACAGCTCTGCTGCTGGCTGCTGCAACAATACAGATGT 5985
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnG 544
|||||
5984 TCCAGCGGCTCTCTACAGTACGCAACTCTCACGACACTAAGGATACAG 5935
544 sTrpThrPheSerTrpLysAsp.....SerAspGluGlyHis 556
|||||
5934 CTGGGCAATGACTTGGTGTATGATACCGCAACACTCCAAACATPAG 5885
557 SerLeuIleAlaAsnTrpThrProLysAsnTrpValProHisProGlu 573
|||||
```

```
222 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSe 248
|||||
207047 GCTCTT.....TCTGGAGATGCCGAGTAACT 207022
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluL 265
|||||
207021 CATATCTGGAGACGACTGTAACCTTCTCAGCAACCAACGACTGTACGTA 206972
265 YSGLYGLVALAAlaIleTyrAlaLysHisMetValLeuArgTyr.....Asn 279
|||||
206971 ATCGCGGACCACTTATATCTTAAGAAGCTTACAGCTGCTTCCCGGGGGGG 206922
280 GlyProValSerPheIleAsnAsn.....SerAlaLysIle 291
|||||
206921 GGGGCTATCTCTCTTCTTAACATATAGTCCAGGTACACGCGAGTAA 206872
291 eGlyGlyAlaAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaG 308
|||||
206871 TCGTCGACCACTTCTATACGCGACGACGACGACGCTAGCTCTTCCAGCAG 206822
308 YLGLGLYSerValLeuPheGlnAsnAsnSer...GlnArgThrSerasp 323
|||||
206821 AAGCAGGAGACATTTACTTCAATGGGAATGCCATTCTTGCACTACACCA 206772
324 GlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSe 340
|||||
206771 CAACTACGAAAGAAATTTCTATTGACATAGATCTACTGCAAGATACAC 206722
340 rSerLeuAlaAlaArgAsnGly...AspIleLeuPhePheAspProIleV 356
|||||
206721 GAATTTAGCTGCAATATCTCGGACATAGCATCTTTTTCACGATCCGAAAT 206672
356 aGlnGlnSerSerSerLysGlySerProLeuProSerSerLeuGlnAla 372
|||||
206671 CTGCTAATACCGCTCGGATTTCTACAGATCTTAAATCTCAATAAGCT 206622
373 SerValThrSerProThrProAlaThrAlaSerProLeuValIleGlnTh 389
|||||
206621 GATCGACGCTAAT.....AGTAC 206605
389 rSerIleAsnArgSerValIlePheSerSerGluArgLeuSerGluGluC 406
|||||
206604 AGATTTATAGTGGGCTGATTTTTCGCGAAAGCTCTCTGAAAGTAC 206555
406 Lu...LysThrProAspAsnLeuThrSerGlnLeuGlnAlaProIleGlu 421
|||||
206554 AAGCAAAAGCTTCAGACACAACCTCACTTACGCTGAACAGACGCTTAACT 206505
422 LeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaPr 438
|||||
206504 CTAACCTGACGAAATTTAGTACTTAACGTGTGCTCACTCTCGATGACAA 206455
438 oSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrS 455
|||||
206454 AGGCTTTACTGACAGACCCCGGCTCTCTGTATATATGATCGCGGACAA 206405
455 eLeuLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIlePro 470
|||||
206404 CCTTAAAGCAAGTACAGACAGAGCTTAAACAGCTTTTCCATTTCTCT 206355
471 LeuHisSerLeuAspThrGluLysSerValThrIleHisAlaPro.... 485
|||||
206354 GTAGACTCTTAAAGGAGGTAAAGAAAGTGTAAATGCTGCTTCGACGC 206305
486 .....AsnLeuSerIleGlnLysIlePheLeuSerAsnSerG 498
|||||
206304 AAGTAAAAATGTACCCCTTACTGCTCGATTTCTTTTGATAAACCAAG 206255
498 IYAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsn 514
|||||
206254 GC.....AAIGCTTATGAAAAATCACGAC...TTAGCAAAAACCTACAGC 206214

515 .....AsnIleProLeuThrLeuSerLysGluGlnSerI 527
|||||
206213 TTTCATTTGACAGCTCTGCTGCTGTGACTGCAACACTACAGATGT 206164
527 sLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
|||||
206163 TCCACCGGCTTCTACAGTACCAACTCTACGCACTACTATGGTATCAAGCTA 206114
544 sPTPrThrPheSerTrpLysAsp.....SerAspGluGlyHis 556
|||||
206113 CTGGGGAGATGCTTGGGTGTATGATACCGCAAGCAGCTCAAGAAGCTAAG 206064
557 SerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGluArg 573
|||||
206063 ACAGGACATTAGCTGTGACCAATACAGCTACTCTCGAATCTCTGAGCG 206014
573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
|||||
206013 TCAAGACCTTTAGTTCTTAATAGCCTTGGGAGATCTTTTCAGACATCC 205964
590 IAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeu 606
|||||
205963 AAGCATTCAGAGGTGTCATAGAGAGAGAGTCTTGACTCTTTGTTTCAGAT 205914
607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSe 623
|||||
205913 CGAGGCTTCTGGCGCGCGGAGTGCSCAAT...TTCTTAGATAAAGATAA 205867
623 rSerGlyLysProIleAspAsnTrpHisHisArgSerLeuGlyTyrLeuP 640
|||||
205866 GAAAGGGGAAAAA...CCCAATATCCGTCATAAATCTGCTGATGATGTA 205820
640 heGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIa 656
|||||
205819 TCGGAGGTGACAGCGCAACTGTTGTGAAACTTAATTAAGTTTGCCCTT 205770
657 GlnGlnLeuLeuGlyLysSerSerAspSerPheIleThrSerThrGluTh 673
|||||
205769 TGCCAACTCTTGTAGCGGATAAAGATTTCTTAGCGCTAAATAATCAATAC 205720
673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
|||||
205719 TGATACCTATGACAGAGGCTTCTATATCCAAACATTAACAGAAATGTAGTG 205670
688 ..... 688
205669 GGTTCATAGGTGTCTCTTAATAAATCTCTGCTGTGAGATCATAA 205620
689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGlnSerIleHisGluLe 704
|||||
205619 CCCCTCGTTTAGAAGGCGACCTGCTATAGCCACGTCATTAATGATCT 205570
704 uLysThrLysTyrArgSerPheSerLysGlyGlyPheGlySerTrpHis 721
|||||
205569 GAAGCAAAAGTATCTGCGTAT...CCYAGGTGAAGAAAGTCTTGGGCGA 205523
721 eValAlaValSerGlyGluValCysAlaSerIleProIleValSerAsn 737
|||||
205522 ATATGCTTTTAACATGATGTGGAGCTTCTTCTCATTTTAATCTGAA 205473
738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPh 754
|||||
205472 TACGTGATGCTTTTGTGATCACTATGCTCATACATCAAGATTCGTAC 205423
754 eSerGlyThrGlnAspGlyPheGluGlnSerSerGlyGluIleArgSerP 771
|||||
205422 CTATATACGTGACAGACACTTCTCGGAAAGGTACAAAGAAAGATCTT 205373
771 heSerAlaSerSerPheArgAsnIleSerLeuProIleGlyIleThrPhe 787
|||||
205372 TTGATGACAGCAACCTCTTCAATTTATCTTGCCATRAGGGGTGAAGTTT 205323
788 GluLysLysSerGlnLysThrArgThrTyrTyrPheLeuGlnAlaTyr 804
```



```
/transl_table=11
/product="outer membrane protein 5"
/protein_id="CAB37071.1"
/db_xref="GI:4455886"
/db_xref="SPTREMBL:O9S6P2"
/translation="MKSQPSWLVLSTLACFTSCSTVFATNENITGSPDSFDCSTGTG
TYTPKNTTGGIDYTLTGDTTLQNLDSALTLGCSPDSTESLFGKGLSLFNLKS
SAEGALVSTTDKNSLTGFSSTFLAASVITTPSGGAVKCGDLTFPDNNTLFL
KQDYCEENGCAISTKNSLTKNSTGSLSPFGNKSATKGGKALCATGTVDITNNAPL
LFSNNIAEANGCAINSTGCTLTGNTSLVFSNSVATNGGALISGDVADYISGQSL
VFPSGQAVANGCAIYAKKLTLAGSGCGGISFSSNNIVQTTGANGCALISLAAGCSL
SVAQDITFNGNAIVATTPQTTKRNISIDIGSTAKITNLNLAISGHSIFEDYPTANPA
DSTDTLNLKAKADANGSTDSVSGSIVFSGELSDSEKAVANLSTIKOPATLNGNLV
KRGVTLDTKGFTGTAGSSVIMDAGTTLKASTEEVLTGSLGIVDSIGEBKKVIAASA
ASKNVALSGPILLDNOGNAYENHDLGCTODESPYOLASLGATTTDPAVPATVATP
HYGCGTGWCTWVDPAASPRTKTAATLAWTNKGYLDPNPRGCPVAPNSLWSPSDIOA
IOGVIFRSALITLCSDRGFWAAGVAVANFLDOKKGERKKYHKKSGCIAIGCAQOTCSNL
ISFARQQLFSQKDFLVAAKNHDTYAGAFYIOHTIECSGFIQCLDLKLGSHKPYA
LEGOLAVSHVSDNLKTKYAYPEVKSGMGNNAFMNLGASISYEYELHCPDTYAPYI
KLNTLYIRQDSSEKTEGSRFSDSNLFLSLPIGKPEKFSQDNDFSDLTLSVYPD
LIRNDPKCTTALVISCASMEYVANNILARQALQVRAAGSHVAFSPMEFVLQGFVEVACS
SRIVYDLGCKRPF"
complement(11313..11372)
/gene="omp5"
/1535..14321
/gene="omp4"
/1535..11588
/gene="omp4"
/1535..14321
/gene="omp4"
/codon_start=1
/transl_table=11
/product="outer membrane protein 4"
/protein_id="CAB37072.1"
/db_xref="GI:4455887"
/db_xref="SPTREMBL:O86164"
/translation="MKTSLIPWLVSSVLAFLSCHLOSLANELLSPDSEFNGNIDSGTF
TEKTSATYSLTGDFEYEPKGTPLSDSCFQOTDNLFLPLNGSLTGFDFAGHA
GAASTATANKNITFSGFSLSPDSSPTVTTCGSLAGCVNLENIKLVAVAGS
PADGAIKICASFLTGTSGLDALFSNNSSTKGCALATVAGIARVNNVSNAS
VSGCATDDECTSLISNNKFLYPEGNAKTTGCAINPKRSQPELISNNKLLIASN
VAETSGAIHAKKLLSSGCTFEFLNNYSATPKGAISIDAGSELSISVETGNTLN
VAETLTGTSTDPKRNAINISGNEFELRAKKNITFEYDPISEGTSSVLEKLTNN
GSAAGLNPOGTILFSGERTLADDELKVAONLKSFPQPLSGGKLLLOKGVLESTG
FSQEGSLGMDSGTTLSTTAGSITITNGIIVNDSIGLQPVSLAKGASNNVYVSK
LNLIDEGNIESHMFSDOLPSLLKITVDADVDPNVDISSLIPVADPNSEYQSG
QNNVATTPPTATNTKEATPTWTKTGVPSPERKSALVCYVILNGVTFISLOQIYEIG
AIGMEHKQCFWVSMNLEHKTGDENRKGFRTSGVIGSAHTPKDLFTFARFCHL
FARDDKCFIAHNNSTRYGGTLEFKHSHTLQPNQYLRGLAKSESASIEKFPPEPLAL
DVOVSFSDNMRMETHYTSLEPESGSMNECIAGIGIDLPEVLSNPHLFTFTFIPOM
KVENYVSONSFESSSDGRGFSIGRLNLSIPVGAKFVQGDIGDSYVDLDSGFPSVD
VYRNDPOSTATLVMSPDSWKIRGCNLSRQAFILRGSNNVYVNSNCELFCHYAMELRCS
SRNVVDVGTCLRF"
/1579..14644
/gene="omp13"
/14579..16123
/gene="omp13"
/14579..16123
/gene="omp13"
/codon_start=1
/transl_table=11
/protein_id="CAB37073.1"
/db_xref="GI:4455888"
/db_xref="SPTREMBL:O9Z3D6"
/translation="MTILRNLTSCALFLAFAAAOVYVLIHSDGYNCAIINKSLEPK
ITCYPEGTSYIFLIDVRISNVKHDQEDACVFINRGNLSLFEMKNRNFHTNTEEGF
AIIISNRVGDITLISNFSLAFTSAPLPQGGASVILFENRNSSEENVTFCGNYSW
SGAAIYTPYLLGSKASRPSVNLGNRYLVEFRNVNGYGAISTINLLITTGSPSFE
NNHAAHDVNSNGCAIAIARGCSISISVKGDDLIFGNTASOQGNITLHNSIHOSGOF
KULRAVSESGYFYDPISSHSESKITTDLYINAPCKTYEGTISFGCLDIDHEYCAE
NITSTILQVYTLACGTLISDSVYTLQHSFRKDEASTLTMSGCTLLCGCDRVONLH
IILEDIDNEVPVIRAREDKDALVSLKLVAREAVASVYDFQKFAETIPLLELLGP
```

```
alignment_scores:
  Quality: 883.00      Length: 942
  Ratio: 1.541      Gaps: 31
  Percent Similarity: 60.828      Percent Identity: 30.573

alignment_block:
  us-09-677-752-4 x CPN133034/rev ..

Align seq 1/1 to reverse of: CPN133034 from: 1 to: 17280

12  CysAlaIaIaIeLeuSerSerThrAlaIaIaLeuPheGlyGlnAspProLe 28
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11315  TGTTCACACTGTTTTCGTGCAACTGCGAATAATATAGCCCTCTGATAG 11266
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28  u.....GlyGlnThrAlaLeuLeuThrLysAspProAsnHisAlaValC 43
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11265  CTTTGACGGAAGTACTAACACAGCCACTTACTCTCAAAAAATGACACTA 11216
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
43  ys.....ThrPheGlnAspCysThrMetGluSerLeu... 54
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11215  CTGCAATAGACTATACCTCTGACAGACATATAACTCTGCAAAACCTTGGC 11166
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
55  ...PheProAlaLeuCysAlaHisAlaSerGlnAsp.....AspProLe 68
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11165  GATTGCGCAGCTTAAACGAAGGCTTTTCTGACACTACGCAATCTT 11116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68  uTyValLeuGlyAsnSerTyrcySTTPheValSerLysLeuHisLiet 85
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11115  AACCTTGCCGGTAAGGGTACTCA.....CTTCTTTTAAATATTA 11072
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85  hrAspProLysGluAlaLeuPheLysGluLysGlyAspLeuSerIleGln 101
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11071  ACCTACTACGCGAAGGC.....GCAGCACTTTCGTGTACA 11037
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102  AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11036  ACTGATAAATCTGTCGTACACAGATTTTCAGCTTACTTCTTAC 10987
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116  rSerProSerIleLeu.....HisGln 124
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10986  GCGCCCATCATCGGTAATCACACACCCCTCAGAGAAAGCTCACTTAAT 10937
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124  yAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10936  GTGAGAGGAGATCTTACATTGATTAACGAATGAACTATTATTATTAACAA 10887
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141  AsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLe 157
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10886  GATTACTGTGAGAAATGCGGAGCATTTCTACCAAGATCTTCTTT 10837
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157  uGlnHisAsnTyrlleuPheThrAlaPheGluGluAsnSerSerLysGly 173
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10836  GAAAAACAGCACGGGATCGATTTCTTTTGAAGGAATAAATCGACGCCAA 10787
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
174  .....AsnGlyAlaIleGlnAlaGln...ThrPheSerLeuSer 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10786  CAGGAAAAAAGTGGCGCTATTTCGTCTACTGCTACTGATGATATATACA 10737
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
187  ArgAsnValSerProLleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10736  AATAATACGCTCTTACCTCTCTCGAACAAATATCTGACAGCTGACGAG 10687
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203  yGlyAlaIle...CysCysSerAsnLeuIleCysSerCylAsnValAsn 219
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10686  TCGACCTTAATAATGACACAGGAACCTGTACAAATTAACAGAAATACGCT 10638
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219  rGluPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10637  ..CTTGATTTTTCGAAATATGCTGACACCGACCGACGAAATGCGAGCA 10590
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232  AlaIleCysCylIleSerAspLeuAsnThrSerGluLysGlySerLeuSe 248
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10589  GCTCTT.....TCTGAGATGATCCGATGTTTAC 10564
```


788 GluLysSerGlnLysThrArgThrTyrTyrThrPheLeuGlyAlaTyr 804
 2509 GAGAAAGTCTGTGATGTAATGACTTTCTATGATGATGACTTATCCTA 2558
 804 rllleuLaspLeuLysArgAspValGluSerGlyProValValLeuLeu 821
 2559 TGTTCCTGATCTTATCCGCAATGATCCCAATGCACACTACACTTGTAA 2608
 821 ysasAlaValSerTrpAspAlaProMetAlaAsnLeuAspSerArgAla 837
 2609 TCACCGGAGCTCTTGGCAACTATGCCATACCTTAGCACGACAGCC 2658
 838 TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeu...GlnThr 853
 2659 TTGCAGAGCGCTGCAGGACGACTACGCGCTTCTCTCTATGTTGAAGT 2708
 853 rleuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTyrSerL 870
 2709 GCCTGCCGAGTTTCTCTTGAAGTTCGCGATCCACGAGTATTAATATG 2758
 870 euAspLeuGlyThrThrArgPhe 878
 2759 TACATCTTGGCGGCTACCTCCCAATTC 2784
 seq_name: gb_bai:AE002192
 seq_documentation_block:
 LOCUS AE002192 12676 bp DNA BCT 30-MAY-2000
 DEFINITION Chlamydomophila pneumoniae AR39, section 25 of 94 of the complete genome.
 ACCESSION AE002192 AE002161
 VERSION AE002192.2 GI:8163405
 KEYWORDS
 SOURCE Chlamydomophila pneumoniae AR39.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
 REFERENCE 1 (bases 1 to 12676)
 Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uitterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39
 Nucleic Acids Res. 28 (6), 1397-1406 (2000)
 JOURNAL MEDLINE 20150255
 PUBMED 10684935
 REFERENCE 2 (bases 1 to 12676)
 Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uitterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
 Direct Submission
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 On Jun 1, 2000 this sequence version replaced gi:7189226.
 FEATURES
 source
 1..12676
 /location/Qualifiers
 1..12676
 /organism="Chlamydomophila pneumoniae AR39"
 /strain="AR39"
 /db_xref="taxon:115711"
 /note="synonym: Chlamydia pneumoniae AR39"
 complement(90..185)
 /gene="CP0300"
 complement(90..185)
 /gene="CP0300"
 /note="hypothetical protein: identified by glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"

gene
 CDS
 /protein_id="AAF38157.1"
 /db_xref="GI:7189227"
 /translation="MPTSLGRSAIFVYTGSLKHLRLGELEIFOK"
 complement(275..1819)
 /gene="CP0301"
 complement(275..1819)
 /gene="CP0301"
 /note="similar to GP:4376735; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="polymorphic membrane protein G family"
 /protein_id="AAF38158.1"
 /db_xref="GI:7189228"
 /translation="MTILRPLTCSALFLALAAAOVYLHESDGYNCAINNKSLKPGITCEPGSTVITLDVIRISNKKHDOEDAVFLNRSGNLFPMNRRNFTPHNMTSGFCAIISNRVGGTITLTSFSLATSAPLDGGGATYSLCSVINESEETFGCNISWSCAATPYPLLSKASRPVNLISGRVLYFRONVSOVGALISTHNLTLTTGGSFNMHAYHDVNSGCAIALAPGGSISISGKGLDIPFGNVAASODQNTIHNLSILOSAOFKULRAVESGVYFVPISSHESHKITDLYINAPKEKEYEGTISFGCLDHEVCAENLSTILODVLGAGTSLSDGVTLOLHSEFROEASTLTPMSPTLLCSGDARVONLILIEDTNFVPIRAEDKDALVSLKELKVAEEAVSYVDPPQKEAPITLLELGPSPDSLIGETTLERTQVTTENDAVRGFMSLSEETPPSLDKRRITPKKIVFLTWNPETITP"
 complement(2077..4926)
 /gene="CP0302"
 complement(2077..4926)
 /gene="CP0302"
 /note="similar to GP:4376733; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="polymorphic membrane protein G family"
 /protein_id="AAF38159.1"
 /db_xref="GI:7189229"
 /translation="MILRECYIDANSLSKNSITMKTISIPVLYSVSLAFSCHLQSLANEELSPDSFNGNIDSGFTPKTSATYSLTGDFEYFEPKQPI.LSDSCFKQRTDNLWPLGNGHSITFGFIDAGTAGAASATYANKMLTGFSLSLFSDSPSTVTYTGGLTSAGAGVNLNIRKLVVAGNFTADGAILGASFFLTGTGDLFENSSSTGGAATTAGARANTNGVRLSNASTSGAIDDEGTSISNNKLEFECNAKTTGGACATKASGSELIISNNKTLIFASNAVERSGAIIHKKALSLSSGCTEPLRNVSATPRKGATSIDSGELISAETGNTFYRNLTTGTSQVDPFRNAINSGKFPLELAANNHTIFFYVITSEGLSSDVLKINNSAGALNPQCTLLFSGCTLLADELKVADNLKSSFTQPVLSGCKILLQKGVLTSESTSFSDGSLIDSGSTLTGTAGSTITNVLGINVSLGLKOPVSLTAKGASNKYVSGKLNLDENIYESHMSHDLSELSLKITVDADPNVLDISILVPAEDPNSFEGOGQNMVMTDTATNKATATKTFGVSPSRKSLVCMTNGVFTDIDSLQOLVLEIGATGEMHKGFPVSSMTLHKTGDENGKGRFHTSGCYVITGSAHTKDDLFTRFCHLTPARDKCTIANNSTTTCGLTPFRHSHTLOQNTLRLGRAKFSESAIEKFPREIPLALDVOVSFSHSDRMERHYSLSPESSGSMSECTACIGLDLPVLSNPHLPFTFIPQMKVEMVYVQNSFFSSSDGRGCFISIGRLINLSIPYGAKEFQGDIGDSYTYDLSGFVSDYRVNRPSTATLVMSPDWKIRGNLSIQAVLRLGNSNIVVYNSNCLFGHYAMELGRSSRNINVDGTRLR"
 5026..7812
 /gene="CP0303"
 5026..7812
 /gene="CP0303"
 /note="similar to GP:4376729; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="polymorphic membrane protein G family"
 /protein_id="AAF38160.1"
 /db_xref="GI:7189230"
 /translation="MKSQSMVLVLSLACFTSCSVFAETAENTGSPSPGSGNTGTYPKNTTGIDVITLGDITLONLDSALTLGCSBDTRESISPRAGKISLFLIKSSAEGALVYTDKNLSLTGSSSLTFLAAPSVTTPPSGKAGACGDLITDNNGLTFLKODCEENGLAISTKNLSKNTGSLISFSGKNSVATPAGNGALSGADVYISGNOSVFFSGNOAVANGAIAIAKLTLLASGGGGSISNNINVOGTAGNGALISIIAAGCSLSAEAGDITFNGAIVATTPQTKRNSIDIGSTAKIVNLAISGHSIFEDYDPTANTALDSTDTLNKKADAGNSDVSISVSGELSRDEAKVADNLSITLKQPTVLAAGNLVLRGVTLDTRKPTQTAGSSVIMDAGTTLKASTEBVTLTGISTIVDSIGBEKKVYIASAASKNALSGPILLDNGNAVENHDLGKTQDPSFVOLSALGATTTDPAVPTVATPT

```

4181 GAGCGCCCAAGCTCTATGCGGGTCTGTATATATATAGCATGAGCA 4230
687 erleu.....Melysileseralaginalacytyras 698
4231 AGTTGATGATCTCAGCGGTTATTATGAGCCTACACAGTGTGTCA 4280
689 .....
4281 CGGTTTCTAAGAGATTCTATATTTTGGATGACAAATTAACCTATWG 4330
698 ngluserlehisguleuLysThyTyraSerSerSerSerLysglug 715
4331 CCACACGGCCACACACATGACACAGCTCTTATACAGACTAT...CCTGAG 4377
715 lypheglSerTrpHisSerValAlaValSerGluValCysAlaSer 731
4378 TGAAGGTTCTTGGGTATGATACCTGGGCTTACTTGTCTACTGAC 4427
732 lileprolevalSerasnlyserGlyLeuPheSerSerPheSerileph 748
4428 GACCTATCCCGGTATWTAGTCTCTCTCTCTCTCTCTCTCTCTCTCT 4477
748 eserLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGlnGluSer 765
4478 TGCAAAATTAACAGTTGCTCTATGCGCACCAAGATGACTTTAAGAACCA 4527
765 erGlyGluLeuArgSerPheSerAlaSerSerPheArgAsnLysSerleu 781
4528 CACAGAGAGCGCGGCTTTTGAGACAGCATCTTCTACAGCTTTCTGTA 4577
782 proilegylileThrPheGluLysLysSerGlnLysThrArgThrTyT 798
4578 CCTATAGCTATAAATTTGACAACTCTCTCTCTCTCTCTCTCTCTCTCT 4627
798 rTyPheLeuGlyAlaTyrlleGlnAspLeuLysArgAspValGluSer 815
4628 TGATCTTACATGATGATATACCTGATGTGTACCGTCATATATCCAACT 4677
815 lypProvalleuLeuLysAsnAlaValSerTrpAspAlaProMetAla 831
4678 GTATGACAGGATTGGCATCAATGACGTTCTGTTACCAACAGCTACG 4727
832 AsnLeuAspSerArgAlaTyrllePheArgLeuThrAsnGlnArgAla 848
4728 AATCTCTCTAGCAAGCTTTCATAGTTCGCGCGGTACCAATTTGCCTT 4777
848 u...HisArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGly 864
4778 AACCTCTGCTGTGAGATGTTCACTGCTTGGTTTCAATTACCAAGCT 4827
864 lSerHisSerTySerLeuAspLeuGlyThrThrTyraGlyPhe 878
4828 CTTCAAGAAATTATACGTAGTCTTGGCGGTACAGTGGCGGTTG 4871

seq_name: gb_patl:A81829
seq_documentation_block: 2815 bp DNA PAT 21-JAN-2000
LOCUS A81829
DEFINITION Sequence 3 from Patent WO98589953.
VERSION A81829
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2815)
AUTHORS
TITLE
JOURNAL
FEATURES
source
1..2815

```

```

BASE COUNT 790 a 575 c 626 g 824 t
ORIGIN
alignment_scores:
  Quality: 883.00 Length: 942
  Ratio: 1.541 Gaps: 31
  Percent Similarity: 60.828 Percent Identity: 30.573
alignment_block:
  US-09-677-752-4 x A81829
Align seg 1/1 to: A81829 from: 1 to: 2815
12 CysAlaAlaIleLeuSerSerThrAlaValLeuPheGlnAspProle 28
58 TGTTCACAGTGTCTTGTGCAACTGTGAAATATAGCCCTCTGATWG 107
28 u.....GlyGluThrAlaLeuLeuThrLysAsnProAsnHisValAlC 43
108 CTTGACGGAAGTACTTAACACAGGCACCTATCTCTTAATAATACGACTA 157
43 ys.....ThrpheGlnAspCysThrMetGluSerleu... 54
158 CTGGAATAGACTATACCTCTGACAGAGATATACCTGCAAAACCTGGG 207
55 ..PheProAlaLeuCysAlaHisAlaSerGlnAsp....AspProle 68
208 GATTCGGGAGCTTTAACGAGGGGTGTTTCTGACACTAGGAACTTT 257
68 uTyValleuGlyAsnSerTyrcystrpPheValSerLysLeuHisile 85
258 AAGCTTCCCGGTACGGGTACTCA....CTTCTTTTAAATATTA 301
85 hrAspProLysGluAlaLeuPheLysGlyLysGlyAspLeuSerileGln 101
302 ACTCTAGTCTCAAGGC.....CGACACTTCTGTTPACA 336
102 AsnPheArgPheLeuSerPheThrAspCysSerSerLysGlu.....Se 116
337 ACTGATAAATCTGTCTGCTAACAGGATTTCTGAGTCTTACTTCTTACA 386
116 rSerProserilelle.....HisGlnL 124
387 GGGCCATCATCGGTATACACACCCCTCAGAAAGGTCACTTAAT 436
124 ysAsnGlyGlnLeuSerLeuArgAsnAsnGlySerMetSerPheCysArg 140
437 GTGACGGGATCTTAACATTGATTAACATGCAACTATTTTAAACAA 486
141 AsnHisAlaGlnGlySerGlyAlaIleSerAlaAspAlaPheSerle 157
487 GATTACTGTAGGAAATAGGGGAGCCATTCTTACCAAGAACTTTCTT 536
157 uGlnHisAsnTyLeuPheThrAlaPheGluGluAsnSerSerLysGly. 173
537 GAAAAACACACGGGATCGATTCTTTGAAGGATATATACGACGCA 586
174 .....AsnGlyAlaIleGlnAlaGln...ThrpheSerleuSer 186
587 CAGGAAAAACGTCGGCTATTGTGCTACTGTAAGTGAATATATACA 636
187 ArgAsnValSerProLysPheAlaArgAsnArgAlaAspLeuAsnGly 203
637 AATAATAGGCTCTTACCTCTCTCGAACATATTTGTAAGCTGACG 686
203 yGlyAlaIle...CysCysSerAsnLeuIleCysSerGlyAsnValAsn 219
687 TGGAGCTTAATATAGCAGAGAACTGTACAAATTAACAGGAATACGCT 735
219 roleuPhePheThrGlyAsnSerAlaThr.....AsnGlyGly 231

```

684LeuAlaThrSerLeuMetlys..... 690
 2174 TGCCCTCATTTGGAGTATACACGACATGCGAAGATATTCCTCTCAT 2223
 691 11SerAlaGlnAlaCysTyrAsnGlnSer11eHisGluLeuLysThrLy 707
 2224 TTGAATGCCAGCTAACTACAGCTACACTAAATGATGATGATGATG 2273
 707 sTyrArgSerPheSerLysGluGlyPheGlySerThrPheLysAlaLay 724
 2274 CTAACTCTCTAT...CCTGAAGCTCAAGGTTCTGACCAATAATATCTG 2320
 724 alSerGlyGluValCysAlaSerLeuProle...ValSerAsnGlySer 739
 2321 GGGCTCTAGAGCTCGAGAGTCTGCGCTATATCTCCCTAAAGAC 2370
 740 GlyLeuPheSerSerPheSer11ePheSerLysLeuGlnGlyPheSer 756
 2371 CCCTCTCTCCAGGATATTCCTCTTAAAGTTCCAGGACGCTCTACAG 2420
 756 yThrGlnAspGlyPheGluGlnSerSerGlyGluLeuArgSerPheSer 773
 2421 CCGCCAAACAACTTAAAGAGAGTGGCCCTCAAGCCGCTCTTGTGATG 2470
 773 1aSerSerPheArgAsn11eSerLeuProleGly11eThrPheGluLys 789
 2471 ATGAGAGACTAGTGAACGCTCTATCCCTGTCGATTCGGTATGAAAA 2520
 790 LysSerGlnLysThrArgThrTyrTyrPheLeuGlyAlaTyr11eG1 806
 2521 ATCTCCGAGAGTGAAGAAAAATATTCGAGATTCTCTCAACCAATTCG 2570
 806 nasPheLysArgAspValGluSerGlyProValValLeuLeuLysAsn 823
 2571 TGATGCTGATGTAATAATCCCTCTGCGGATCTCTTAAGTCTGACG 2620
 823 1aValSerThrPaspAlaPheMetAlaAsnLeuAspSerArgAlaTyrMet 839
 2621 GAGCTCTTGGAGCTTCGCTATGTAATAACCTCGCAGCAGACCTCTTA 2670
 840 PheArgLeuThrAsnGlnArgAlaLeuHis...ArgLeuGlnThrLeuLe 855
 2671 GCAAGCTGTCGAAAGCCATCTGCTCTCCCTCACTGACAACTCTGCG 2720
 855 uasAlaSerCysValLeuArgGlyGlnSerHisSerLysSerLeuAspL 872
 2721 GGAAGCTCTTATGAGCTTCGCTGAGCAGACACTTACATGATGAT 2770
 872 euGlyThrThrTyrArgPhe 878
 2771 GTGGGCTAAGATACATCTTC 2790
 seq_name: gb-ba2:CP072499
 seq_documentation_block:
 LOCUS CP072499 4926 bp DNA BCT 13-MAY-1999
 DEFINITION Chlamydia abortus putative outer membrane protein gene, partial cds; and putative 98 kda outer membrane protein gene, complete cds.
 ACCSSION U72499.1 GI:1657776
 VERSION 072499.1 GI:1657776
 KEYWORDS Chlamydia abortus.
 SOURCE Chlamydia abortus.
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 4926)
 AUTHORS Longbottom,D., Russell,M., Dunbar,S.M., Jones,G.E. and Herring,A.J.
 TITLE 98kda protein genes from ovine abortion strain S26/3 Chlamydia psittaci
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 4926)
 AUTHORS Longbottom,D.
 TITLE Direct Submission

JOURNAL Submitted (25-SEP-1996) Moredun Research Institute, 408 Gilmerton Road, Edinburgh EH17 7JH, Scotland, UK
 FEATURES
 source Location/Qualifiers
 1.4926
 /organism="Chlamydia abortus"
 /isolate="ovine abortion S26/3"
 /specific_host="Ovis aries"
 /db_xref="taxon:83555"
 /clone="S112"
 /dev_stage="elementary body"
 /note="submitted as Chlamydia psittaci"
 <1.1950
 CDS
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAB18187.1"
 /db_xref="GI:1657777"
 RBS
 /note="putative"
 2083..2089
 FCHGAFELGSSSNVNDIGRRKF
 CDS
 /codon_start=1
 /transl_table=11
 /product="putative 98 kda outer membrane protein"
 /protein_id="AAB18188.1"
 /db_xref="GI:1657778"
 /translation="MRPSLYKILISSTLTLPISFPHSOLAARVALTOESIIIDANCAFS
 POSTRACGTIYNVESDISIVDVGOTAAALASSAPOTANLTFKGNHSLSTNNAG
 NPAGIYNVTRADRIITLIDPSKLSRECEPSSLVNTGKAMKSGALNLANNSIIFSD
 NISAENGAIISCAKASLTGSSKEISFTTNSIAKGAIAIAIGAILSDMOGTIRFSG
 TAVNSGAVYSEASMTIACNNHVAFAFNNAVSSSDCGAIICSKTGSAPLTIRDNK
 VLFEENGTSSAKGAIYTDKILITSGCPAFINNKYTHATPAGKAIAGIAANGCSLTA
 EHGDIITFDDNNLWATODNATIKRNAINIEBNGFVNLRASGKGTISPYDITVEGNAD
 LITINKAREDKTYNGRIITFSGEKLEPEOAAVANDNLTPTDPTITLAAAGLVRSGEV
 EAKTVYVOTGSLIMDAGTKLSAKTEDATLTLDAIPNLTLDGKRAVDVAAGAKNVT
 LSGAIGVIDPTSGEYENHRLNDTLALGIGIOLSGKGSVTTNPSHVGALEHYGOG
 MNSVSWKXNDNSDPKPTOTA1FTWNTKGYVPNDRAPVLNLNLSMGSFIDRLSIOVLE
 RVSYDSLLETRGCLMVSICGNPFHDKDANERKFRHISGTVLGATNTSRSDLSVAF
 COLFAKDXLYLVSKNAANYAGSVYOHVSKRFDLITRLFNGPNPCCSGSKRILPFLD
 AOTYCHTANNMNTSYTDPEVKGSGNNTLILITLSTYPTIPYFSSITSDIAPFAKL
 QVYVHAQDDPFKEPTTEGRVSESSDLNLSVPIGIRFELKLSVGRSAYDITLWYIPDY
 RHNSPCTGLAINDVSWLTATNLARQAFIVRAGNIHALTSGVEMFSQFELRSSSR
 NYNVDLGARVAF"
 4884..4905
 /note="potential 9-base terminator sequence"
 stem_loop
 BASE COUNT 1472 a 1033 c 1031 g 1390 t
 ORIGIN
 alignment_scores:
 Quality: 896.50 Length: 848
 Ratio: 1.731 Gaps: 25
 Percent Similarity: 61.085 Percent Identity: 30.307
 alignment_block:
 US-09-677-752-4 x CP072499 ..
 Align seg 1/1 to: CP072499 from: 1 to: 4926
 98 LeuSer11eGlnAsnPheArgPheLeuSerPheThrAspCysSerSerLy 114
 2442 CTACAGCTGACAGATTTTCTAAGTCTTAAAGCAATGCCCATCTC.. 2489

SOURCE Chlamydomophila pneumoniae CWL029.
ORGANISM Chlamydomophila pneumoniae CWL029
REFERENCE 1 (bases 1 to 15068)
Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia. Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S. Comparative genomes of Chlamydia pneumoniae and C. trachomatis Nat. Genet. 21 (4), 385-389 (1999)
TITLE 99206606
JOURNAL
MEDLINE
PUBMED 10192388
REFERENCE 2 (bases 1 to 15068)
Kaiman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source
1. 15068
/organism="Chlamydomophila pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
99. 626
/gene="Cpn0439"
99. 626
/gene="Cpn0439"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18583.1"
/db_xref="GI:4376722"
/transl_table="MSSTLNGVFPSPSLPEESADLTNNKEIVLAGEKNVFTLHSPIM HIAITLIVLVALAGIATIGCYSOSILLIIVAGIVLTITLLCOALVGFKEFIROL POOLHTTVQVIFREKIRPESSIDLVINAOKRTTODITKLIEELCDLSQKFKIQSITLYO KKELSHKNEKTNN"
651. 1289
/gene="Cpn0440"
651. 1289
/gene="Cpn0440"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18584.1"
/db_xref="GI:4376723"
/transl_table="MATSVAPEPSPSSPLSHATEVLNPNAYITQHPIDPAPEMTEF HSKLSRHLKALPILLLIGCTISAGVAGYGNMIGIGIGIIVLTILALLAIPL LKNGCTKLIDEISODISSISCPQWYGLMPESTKSHLPEIPLTONEKRIILNEI EKKESIQWLEKITECONKLAKOPKRSKSSKSTKSLKSNPVLIPDC"
1437. 2387
/gene="Cpn0441"
1437. 2387
/gene="Cpn0441"
/codon_start=1
/transl_table=11
/product="CT007 hypothetical protein"
/protein_id="AAD18585.1"
/db_xref="GI:4376724"
/transl_table="MFKLFIHIAFAFGHVLNPIPIVODAGIDEACKNPPRPESA OYOLKVNDAKREKLPHORIGYROYDGFCTLPITPHSCILLFSTCYICADIDKWSL PISFDPNGLGKATFODTISFYNYVLISICATYLSLKNGQMSILSLVUPKINEMGYS LVQVLSGKYGATEKLSAIFGVINTEGLOHOKRAMPVGVSYKATDQTLNCTIVPVS IDYRSTVCNGLAYRLTFRRKRLKLNHLISSRGIFEYOGREIEANVVKLTPPGSEIK GFYSGICNDISIDADHNNKTSHTFKTSAFGGSAYVMNF"
2559. 3077
/gene="Cpn0442"
2559. 3077
/gene="Cpn0442"
/codon_start=1
/transl_table=11
/product="CT006 hypothetical protein"
/protein_id="AAD18586.1"
/db_xref="GI:4376725"
/transl_table="MGEFNICKOGSOLYNGIPPERIILARKLNCKAKSYPRALITIEV LVSSVIGALKVILIPCASTYVALTLPLRALFNAIKTKSCQHLASYAMAMLHLITLIV

gene
3275. 4528
/gene="Cpn0443"
3275. 4528
/gene="Cpn0443"
/codon_start=1
/transl_table=11
/product="CT005 hypothetical protein"
/protein_id="AAD18587.1"
/db_xref="GI:4376726"
/transl_table="MSQPPINPLGQPPVPAAPSPGQSPVYKRLKTSSTGLFKRPIIT PDKYPMKRYVDTGIILALAIILSLILLTASGNSLMLYALAPALAGLGYLLISDI LIDSPKARKGEAITAIVPIYLVAIAGIACAFVASSGIMVIFANPMFVGLITVGL YFMSLNKLTIDVFRREHLLRMEKTOFTAEPIILVPSADDAKKIIVERKVDLSARM EHEKASODAHARRICREACOSPFYSSRNPRKRSFGLSKFKPDSASTASAI SPFRKDDQPTHEKDLRSSFGSGASAFPIPMPSSPSPNSTGTIVLHPEVYPRKG KEPSIRVSSSSRSRPRDQDKQ0Q0Q0DEEQKQKRSKSKSNQSLTTPPDGKST ANLSPSPFSSDEREKRRKRNK"
4833. 9056
/gene="pmp_6"
/note="Cpn0444"
4833. 9056
/gene="pmp_6"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G/I Family"
/protein_id="AAD18588.1"
/db_xref="GI:4376727"
/transl_table="MKYSLPWLITSSALVFSLHPILMAANTDLSSDNYENGSGSAF TAKESVDSAGTYTTLSDVSIINVAIIPADKSCFTNNGALSEFGADHSILQTLAL THDGAIINNTNALSFEFGSSLLIDSAFATGSGGKGAICVNTGATFEDNASTV LOKNTSEKMGIAVSAVSIIDLAKTITTAALIDONTSTKNGALCSTANTTVQSGSVTF SNTWATDKGCGIYSKEKESQIDANTGCVVTFKNTAKTGGASNDONALITGTOVLYO ENKTTGSAOANNPBGCGGATCCYLATATDKGLAISQNEHSFTSNTTANGCAIYA TKCTLDGNTLTFTDONTATAGCGGAIYETEDFSLKSGTGTFTSTNKTGALYSK GNSSLGTNNLLESGNKATPSSNSANEGCGGALIAFIDSGVADKGLSTANNQEV SLTSNAATVSGGAIYATKCTLTGNSGLTFEDGATAGSGAIYETEDFTLSTGTV FSTNTAKNGALYSKGNNSLSCNTNLLFSGNKAQPSNSANEGCGGALIAFISAS VSTRKGIAMEDNENVSLSGNTATVSGGAIYATKCAHGWNTLTPEGNTAEATGGLIY ETEDFTLSTGTVFTSTNKTACALTKKNTSTKKAIFYSNSKATATATTTTDO ECGGAILCINISDIDATKSLTENESLFTNNTRAKRGGGIIYAPKCVISQESINF DONIAETSGGAIYSKRLSTANGPVSFTNSGCGGAIYIAGSGLSLEAIDGDTFS GRARTEGTSIPNSIHLGAGAKITKLAAPGHIYIYDPTMKPAGSGTIEELVYINP VKAIIPPOPKPGPIASGVVPAVPAAPAGHTIYVSGGLTIPDQASIPANTTILNOK IALACGNVATKSGATLOVYSFTQOPDSYFPMAGTILETTTNNNGSDIDKLSLNL DALDKRMTIYIYANSTSGGLKISGLKPLNNKSGSYDNGKANNLNPFLDLSSTST VALDQNPILPSSMAAPDITGIGSWTLVPRVGGGVTLVAEQALGITPKPELRATLV PSLNNAVYNIHSIQEITATMSDAPSHPGIIGIGIYGFHODKQENAGFRLISRGY IIVGSMNTPEOETTFVAFSOLFSGKSDYVSDIKQVYAGSLCAQSSVYIPLHSSLR HVLSKVLPELPGETPLVHGOVSYGRNHNMTTKLANNTOGKSDMSHSFAYEYVGSIL PVDLNYRYLTFSYSPYVKLOVYVNOCKGOEVAADRITDASHLVAVSIPMGLTFKHES AKPPSALLTITGIVANDAYRDHPHCLSLTNGTSWSTPATNLSROAFPAASGLKLLH GLDCEASGSCIEBSSSSRSSTYNANGTRYSE"
9318. 12128
/gene="pmp_7"
/note="Cpn0445"
9318. 12128
/gene="pmp_7"
/codon_start=1
/transl_table=11
/product="Polymorphic Outer Membrane Protein G Family"
/protein_id="AAD18589.1"
/db_xref="GI:4376728"
/transl_table="MKSSVSMFLFSSIPLEFSSLSIYAAEVTLIDSSNNSVDSNGTFTF VESTIDAAAGTIVSLSDVSTFONAGLGIPLASGPLEGGLTTOGNOHAKRFIN AGSSAGTAVASTAADKNLLEFNDPSTRISITISCSLLSPGOCALSVGSLSTGNSOI IPTONFSSDNGVINTKNTFLSGTSQFAFSSNFTGQSGVAVATGITTLENSPGI VFSFONLAKSGCALYSTDNCSITDNFQVYIPGNSAMEAQAQACATCTTDDKIVTL TGNKNSLSTNTNALTLYGCAISGLKVISISGAGTLLTOSNISGSSAOGGAGLISASG ELAISNTSGDITFNNNOVNGCSTSRNANIITDFAKVSIRARATOSGTYFVDPITNPG TAASTDTLNLADANSEIEYGAIFYSEKISPEPEKALAAVNTSTIRQPAYLAGDL VIRDGTVTFKDLTOSPGSRIIMDGGTTLTSAKEANLSLNGLAIVNLSSLDGINKAALKT

Quality: 940.50 Length: 910
Ratio: 1.761 Gaps: 32
Percent Similarity: 58.681 Percent Identity: 31.538

alignment block:
US-09-677-752-4 x AB1839 ..

Align seg 1/1 to: AB1839 from: 1 to: 2787

```
85 ThrspProLysGluAlaLeuPheLysGluLysGluLysLeuSerIleG1 101
|||||.....
154 ACNAATGACATGACATCTATATATCTACAGAGGAGTGTCTCAATAC 203
|||||.....
101 nasPheArPhe...LeuSerPheThrAspCysSerSerLysGluSerS 117
|||||.....
204 CAATGCAAGATCTCCGACAGCTTAACCCCTCTGCTTAACAACACTA 253
|||||.....
117 erProSerIleIleHisGluLysAsnGly...GluLeuSerLeuArGsn 132
|||||.....
254 CTGGGAATCTTTCTTCCAAAGCCAGCTACCAATTTCTCTCAAAAT 303
|||||.....
133 .....AsnGlyse 135
|||||.....
304 ATCGATCGGGAGCGAAGCTGACTTACCAATACAGCTGCAATTAAGCT 353
|||||.....
135 rMetSerPhe.....CysArGsnH 142
|||||.....
354 TCTCTCCTTTGACGATTCCTCTATTGTCATATACAAACCAAGATG 403
|||||.....
142 IsAlaGluGlySerGlyGlyAlaIleSerAlaAspAlaPheSerLeuG1n 158
|||||.....
404 CTACCACAGGAGACAGAGCCATCAAGTCCACAGGAGCTTTCTATTCG 453
|||||.....
159 HisAsnTyLeuPheThrAlaPheGluGluLysAsnSerLysGlyAsnG1 175
|||||.....
454 TCGAAGCTT...AGTGTCTACTTTGGCCAAACCTTTCTATGACAAATG 500
|||||.....
175 yGlyAlaIleGlnAlaGlnThrPheSerLeuSerArGsnValSerProI 192
|||||.....
501 AGCGCCCTCCCAAGCAGCTCTATCAGTCTATCCCTAAACCCCAAC...C 547
|||||.....
192 IeSerPheAlaArGsnArGsnAlaAspLeuAsnGlyGlyAlaIleCysCys 208
|||||.....
548 TACGCTTTGCCAAAACAAACGCAACGCAAAAGGGGGTCCCTATTC 597
|||||.....
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrG1 224
|||||.....
598 ACGGAGGAGATTACATTAACATAACGTTAACTCAGCATCATTTCTGA 647
|||||.....
224 yAsnSerAlaThrAsn..... 229
|||||.....
648 AAATACCCCGCGAACAATGGCGGAGCCATTACAGGAGACTACCACTT 697
|||||.....
229 ..... 229
|||||.....
698 TTATTAGCAGCAACAAGCAATTAGCTTTATTAACAATAGTGTGACCGCA 747
|||||.....
230 .....GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
|||||.....
748 ACCCTGACGACAGAGGAGGAGCCATTACTGTAGTAGT.....ACATCGAGC 791
|||||.....
242 rGluLysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 259
|||||.....
792 CCCCAAACCACTTAACCTATATCAGACACCGGAACTGAACCTTATAG 841
|||||.....
259 erAsnSerAlaLysGluLysGlyAlaIleTyAlaAlaLysHisMeVal 275
|||||.....
842 GAAATACAGCAATAGTAGTGTGGGGGATTTATACGACAATCTAGCTT 891
|||||.....
276 LeuArGtyrAsnGlyProValSerPheIleAsnAsnSer..... 288
|||||.....
892 CTTTCTTGTGAGAGACTACGCTTTTAAACAACTCTGCTATAGATAC 941
```

```
289 ...AlaLysIleGlyAlaIleAlaIleGlnSerGlyLysLeuS 304
|||||.....
942 TGCAGCTCCCTTAGAGAGACCAATTGCCATTGCTGACTCTGATCTTGA 991
|||||.....
304 erIleLeuAlaGlyLysGlySerValLeuPheGlnAsnAsnSer..... 318
|||||.....
992 GTCTTGGCCTTGTGGAGACACTCTTTGAAGGAAACAGTAGTC 1041
|||||.....
319 ...GlnArGtyrSerAspGlnGlyLeuValArGsnAlaIleTyLeu... 333
|||||.....
1042 AAAGAGACTTCTTGACACTCAGACACTACCGAATTTCTATTAACTACG 1091
|||||.....
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArGsnGlyAsp...I 349
|||||.....
1092 AAACACCAATGCTTAAGATTGACAGCTGCGAGCCTCAAGCAACTACTA 1141
|||||.....
349 IeLeuPhePheAspProIleValGlnLysSerSerLysGluSerPro 365
|||||.....
1142 TCTACTTCTATGATCTTAACCACTACCTACT..... 1176
|||||.....
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
|||||.....
1177 .....CCAGCTCTCTCAGATCTCTAACTTAATG 1208
|||||.....
382 aSerProLeuValIleGlnThrSerAlaAsnArGSerValIlePheSerS 399
|||||.....
1209 TCTGACCTTCCAGGAAATCCGCTATACAAAGAACCTGATTTCTG 1258
|||||.....
399 erGluArGLeuSerGluGluLysThr...ProAsnAsnLeuThrSer 414
|||||.....
1259 GAGAGAGCTTCCGAAAGCAGAACCTGCAAGCTGATTAATCTCAATCT 1308
|||||.....
415 GlnLeuGlnGlnProIleGluLeuLysSerGlyArGLeuValLeuLysAs 431
|||||.....
1309 ACATTTACGACCTCTACCTCTGCGGAGGCACTCTCTTAATC 1358
|||||.....
431 pArGAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
|||||.....
1359 AGCAGTCACTCTACTGCTTAAGCTCTTTCCGCAATCCCGGCTCTACCC 1408
|||||.....
448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
|||||.....
1409 TCTCATGATGAGGAGGAGCCATTAGAAACCGCTGATGGGATCTACT 1458
|||||.....
465 AlArThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
|||||.....
1459 AATATATCTTTGTCATGTAGATTCCTTAAGAAGACCAAGAGCCATAC 1508
|||||.....
481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer. 497
|||||.....
1509 CCTAAAGCAGACACAGCAAGCT...CAGACAGTCACTTATCTGATGCC 1555
|||||.....
498 .....GlyAspGluAsnPheTyrGluAsnValGluLeuLeu 509
|||||.....
1556 TCTCTCTTGTAGATCTTGTGAATGTCTACGAAGATGTC..... 1596
|||||.....
510 SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLy 523
|||||.....
1597 .....TCTTGAATTAACCTCAAGCTTTTCTGTCTCACTTCTTACTGC 1640
|||||.....
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer.... 537
|||||.....
1641 TGAGAGACCCCGCAATATTCACATCAGAGC.....TTACTCTGTATC 1684
|||||.....
538 .....HisPheGlyTyGlnGlyAspTrpThrPhe 547
|||||.....
1685 CCTAGAAAAAATCTATCTCATTTGGGATGCCAAGGGAATTTGGGCATTA 1734
|||||.....
548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
|||||.....
1735 TCTTGCCAGAGATACGCGACATAATCCAAACAGCAGACTCTTACTCG 1784
|||||.....
562 pThrProLysAsnTyValProHisProGluArGlnSerThrLeuVala 579
```

```
gene complement(8391..10808)
/ gene="gyrB.1"
/ complement(8391..10808)
/ gene="gyrB.1"
/ codon_start=1
/ transl_table=11
/ product="DNA gyrase, subunit B"
/ protein_id="BA98485.1"
/ db_xref="GI:8978649"
/ translation="MDPKENYDASAITVLEGLQAVRERPMYIGDTGITGLHLIYVE
VVDNSIDAMAGCSRIDRILEDGIVIVDNGRIPIEVEHRESAKOGREYSALEYV
LTVLHACKKPKDKSYKSGVJLHGVCSVNAISELVATVVEKDKKCOMESRCPVY
PLOYVSVSDROGTEIVFPDPKISTCTCFDUSILMKRLRELAFILNRGITIYFEDRDV
SFDEVTFEYEGGIQFVSYLONKESLFSPLIYICGTAVGDGDI EFPAALQWNSCY
ELVSYANNIPRQGTHTLGTSTALTRVINTYIIVAHNLANKNLATLGEDIREGLTA
VISVKNPPOEGOTKOKLNSDVSAQVYGEALITFEENPQIARMIYDKYVAA
QAREAKKARELTLEKSLADSLRPLGLIDCLENDPEKCEMYIVEGDSAGSAGKGR
RRFOALIPRIRKILINVEKARIQIIONOEIGITIALACCICADQNFNLSKIRYRIL
MTDADVDSHIRTLLTFYRHMTALITNECVYIAOPPLKYVSKKDDRYILSEKMD
SYLLMLGTNESSLEKSTERELRGALLESFINLVLDVESFINLLEKKAIPSEFLEM
KEGIGVPLVYLAIPATMOMGGRYLYSDEKEEALQDETHKRIIELLYVAFAVDIQN
LKEVGLDISYLLPQKNEIVIGNEDSPSCNICYATLEVINLKLGRKGIETORYK
LGENVADQLMDTMNPEQRILHVSLEKDAVEADHIFTMLMEVEVPRREFESHLSI
RINNDI"
complement(10812..11159)
/ gene="CpJ0276"
/ complement(10812..11159)
/ gene="CpJ0276"
/ codon_start=1
/ transl_table=11
/ product="CpJ91 hypothetical protein"
/ protein_id="BA98486.1"
/ db_xref="GI:8978650"
/ translation="MFLKRRKGGSSQVONKRTASPIKHAHYLHNYLOELOKIMARP
HDAIDAMNOVFEDKXKMSQAIIGFRDHLIVKVYNSLYALLKQTPQNDLMSLYOVA
SHVQIREQFLIG"

alignment_scores:
Quality: 947.50 Length: 910
Ratio: 1.771 Gaps: 32
Percent Similarity: 58.791 Percent Identity: 31.648

alignment_block:
US-09-677-752-4 x AP002546 ..
Align seg 1/1 to: AP002546 from: 1 to: 300650

85 thrasprolysglualeuhenlysglyaspserleu101
|||||
|||||
200438 ACAATGCAATGCACTATCTAATCTAACAGGGAGTGTCTCAATCAC 200487
101 nasnpheargphe...leuserpethraspcyserserlysgluser117
|||||
200488 CAATGCAGATCTCCGACACCTCTAACCGCTTCCTTTAAAGAACTA 200537
117 erproserillellehisglinsasnly...gluleuserleuargasn 132
:::
200538 CTGGGATCTTTCTTCCAAAGGCCACGCTACCAATTTCTCTCAAAAT 200587
133 .....asnlyse 135
200588 ATCGATCGGAGCGCAACTACTTACCAATACACCTGCCAAATTAAGCT 200637
135 rmeuserphe.....cysargasn1142
:::
200638 TCTCTCTTTTACAGATTCCTCTATTGTCTACCTAATACAAACCGAATG 200687
142 tsalagluglyserglycyala11eseralaspalapheserleu158
:::
200688 CTACCACAGACGAGGACCATCAAGGCTCAGAGCTTGTCTTCAATTCAG 200737
159 hisasnlyrleuphehrallaphelgluanserSerlysglyasn1175
|||||
```

```
200738 TCGAATAT...AGTGTACTTTGGCCAAACTTTTCTATATGACATG 200784
175 yglualaileglalaglnthrpheserleuserargasnvalserpro1192
|||||
200785 AGCGGCCCTCAAGGAGCGCTCTATCAGTATATCGCTAAACCCCA...C 200831
192 leserphealaatargasnargalaspheusnlyglyala11ecyscs 208
:::
200832 TAACGTTTGGCAAAACAAAGCAACCAACCAAAAGGGGTGCTCTATTCC 200881
209 seraspheuleuilecysserglyasn...valasnproleuphehrarg1224
:::
200882 ACCGAGCGGATTCATTAATTAACCTTAACCTCAGCATCATTTTCTGA 200931
224 yasnseralathrasn..... 229
200932 AATACCAGCGGCAACAATGCGGAGCCATTACAGGAAGTAGCAGTT 200981
229 ..... 229
200982 TTATTAGCAGCAACAAGCAATTAGCTTTATAACAATAGCTGACCGCA 201031
230 .....glyala11ecyscsyaleseraspheusnrhse 242
201032 ACCTCAGCTACAGCGGAGCCATTACTGTAGTAGT....ACATCAGC 201075
242 rglulysglyserleuserleuala1cysasnnglnluthrleuphealas 259
201076 CCCCAACAGCTTAACTCTACAGACACGCGGAGCACTGACTAGTTATAG 201125
259 ernasnseralalysglycyala11etlyrlyalalyshismetval 275
201126 GAATATACAGCAATTAAGTGTGGGGCGATTATACGTGCAATCTAGTT 201175
276 leuargtyrasnlyprovalserpheleasnanser..... 288
201176 CTTTCTTCTGAGAGACCTACGCTTTTAAAAACACTCTCTATAGTAC 201225
289 ....Alaly11eclyglyala11ealaleal1eglnserglyserleus 304
|||||
201226 TGCAGCTCCCTTAGAGAGCAATTCGATTGCTGATCTGATCTTGA 201275
304 erlleleuala1glyglylyservalleupheglnasnanser..... 318
|||||
201276 GTCTTTCGCTCTTGTGTGAGACATCATCTTTGAGGAAACACAGTAGTC 201325
319 ...Glnarghrseraspnglnlyleuvalargasnlatletyrlleu... 333
201326 AAGGAGCTTCTTGATCAGATCAGACACCTACCAAGAAATCTTTAAGATCG 201375
334 .GluLysAspAla11leuserSerLeuGluAlaArgasnGlyAsp...I 349
:::
201376 AAMCACCAGATGCTAAGATGTGTACAGCTGCGAGCCTCTCAAGCAATACTA 201425
349 leuuphepneaspproilleval1cnglnuserSerlysgluserpro 365
||
201426 TCTACTTCTATGATCTTATACA..... 201448
366 leuproserSerLeuGlnAla1aserValThrSerProThrProAlaThrAl 382
:::
201449 .....ACTACATCACTCGACGCTCTCAAGATCCTTAACCTTAATTCG 201492
382 aSerProLeuVal11eglnThrSerAla1asnArgSerVal11lephesers 399
|||||
201493 TCTGACCTTCGAGGGAATCCTGCATATCAAGAACCATGATTTTCTG 201542
399 ergluatrgluserglygluglyluthr...proaspasnleuthrSer 414
:::
201543 GAGGAGAGCTCTGGAGAGCAAGCTCAGAAAGCTGATTAATCTCAATCT 201592
415 GlnLeuGlnInpRollieglnleuLysSerGlyArgLeuValleuLysas 431
:::
201593 ACATTTACGACACTTAACCTTTCGCGGAGGCAAGCTCTCTCTTAATTC 201642
```

415 GluLeuGlnGlnProIleGluLeuLeuLysSerGlyArgLeuValLeuLysAs 431
 5135 ACAATTGACGACCTTAACCTTGTCCGACGGGACCTCTCTTAATAATC 5184
 431 PArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
 5185 AGGAGTCACTAGTACTAGTCTTTTGCATCTCCGGCTCTACCC 5234
 448 euIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
 5235 TCCCATGTGATGCGAGGACCATTAAGAACCCGTGATGGATCATCATC 5284
 465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
 5285 AATAATCTGTCTCAATGTAGATTCCTTAACACAGACCAAGAGGCTAC 5334
 481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
 5335 GCTAAAGACACACACAGCAAGT...CAGACAGTCACTTATCTGATGCG 5381
 498GlyAspGluAsnPhetyrGluAsnValGluLeuLeu 509
 5382 TCTCTCTGTAGATCCTCTGGAATGCTCTAGAAAGATGTC..... 5422
 510 SerLysGluGlnAsnAsp.....IleProLeuLeuThrLeuSerLy 523
 5423TCTTGAATTAACCTCAACTCTTCTTCTCTCACTCTTACTTC 5466
 523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
 5467 TGAGACCCCGCGAATATTCACATCACAGAC.....TTAGCTCTGATC 5510
 538HisPheGlyTyrgInGlyAspTrpThrPhe 547
 5511 CCCTAGAAAAAATCCTATCCATGGGATACCAAGGGAATGGGCATTA 5560
 548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
 5561 TCTTGCCAGAGGACTGCGCACTAATTCGAAGCAGGCACTTACTCTG 5610
 562 pThrProLysAsnTyraValProHisProGluArgGlnSerThrLeuVal 579
 5611 GACAAAAACAGATACATCCGAATCTCGAGCTGCGAAGCTTACTG 5660
 579 IAsnThrLeuTrpAsnThrTyrsSerAspMetGlnAlaValGlnSerMet 595
 5661 CTAAACCCCTATGGGATCTTGTGTGATGTGCGCTCCATCAACAGCTT 5710
 596 IleAsnThrIleAlaHisGlyValAlaTyrrLeuPheGlyThrTrpGlySe 612
 5711 GTAGCCACTAAGTAGCCCAATCTCAAGAACTCGCGCATCTGCTGTGA 5760
 612 rAlaValSerAsnLeuPheTyrrAlaHisAspSerSerGlyLysProIle 629
 5761 AGGAGTCTCGAATCTTCT...CATAAAGATGACAGCAAGAAATA 5804
 629 sPAsnTrpHisArgSerLeuGlyTyrrLeuPheGlyIleSerThrHis 645
 5805 AAGGTTTTCGCGACATAGTGCAGGTATGTGTGAGCGCGCTCAACA 5854
 646 SerLeuAspAspHisSerPheCysLeuAlaAlaGlyGlnLeuLeuGly 662
 5865 TTAGCTTGTGAATCTTATCTACCTGACGCTTCTGCCAATTAATGGGA 5904
 662 sSerSerAspSerPheIleThrSerThrGluThrSerTyrrIleAlaT 679
 5905 AGATAGAGTCACTTATATAATAAATAGAGCTTCTGCTATGACGCT 5954
 679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
 5955 CTCTCATCTCCAGCATCTAGCGACCTTGTCTTCTCAAGCTTGTACGC 6004

691IleSerAlaGlnAl 695
 6005 TACCTTCTGATCTGAAGTGAAGACGACACCTCTCTCTTTATCTCTCAGAT 6054
 695 aCysTyraSnGluSerIleHisGluLeuLysThrLysTyraSerPheS 712
 6055 CAGCTATATCTATAGTAAATACTATGAAGAACCTATACACCAACGAC 6104
 712 erLysGluGlyPheGlySerThrPheSerValAlaValSerGlyGluVal 728
 6105 CAAAG...GGAGACAGCTCGTATATAGACGGTTCGCTGACCTT 6151
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSer 744
 6152 GCGAGCTCCTACACACACTCTTAAAGCATAGAGGCTCTCTCCACCC 6201
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGly 761
 6202 GTATTTCTCTTCTCATCAAGTAGAAGCTGCTGATCATACCAAGATAGCT 6251
 761 heGluGlu...SerSerGlyGluIleArgSerPheSerAlaSerSerPhe 776
 6252 TCAAGAAGACGTAAATACCTTGTAGCATCTTTCGATACGGGATTTA 6301
 777 ArgAsnIleSerLeuProIleGlyIleThrPheGluLysLysSerGlnLy 793
 6302 ATTAACCTCTCTGCTGCTATTGGAATACCTTCGAGAGATTCGAGAAA 6351
 793 sThrArgThrTyrrTyrrPheLeuGlyAlaTyrrIleGlnAspLeuLys 810
 6352 CGAGCTGCGCTTACGAGAGCTACTGATCATCTAGTCCGATGCTATC 6401
 810 rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerTrp 826
 6402 GTAAGATCTGATGACGACGACCTCTCTTAATCAACAATACCTCGTG 6451
 827 AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
 6452 AAACCTACAGACGAATCTCTCAAGCAACCTGTATCGGAAGACGACG 6501
 838TyrrMetPheArgLeuThrAsnGlnAlaLeuHisArgLeuG 852
 6502 GATCTTTATCTCTCTCT.....CCAATCTT 6530
 852 InThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerTrp 868
 6531 AGGTACACAGTACCTATCTATGCAATTCGTGGATCTTCACGCACTAC 6580
 869 SerLeuAspLeuGlyThrThrTyraArgPhe 878
 6581 AATGCAGATCTTGAGGCTAATCTCCACTTC 6610
 seq_name: gb_ba2:AP002546
 seq_documentation_block:
 LOCUS AP002546 300650 bp DNA BCT 08-DEC-2000
 DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
 section 2/4.
 ACCESSION AP002546 AB033782 AB033783 AB033784 AB033785 AB033800 AB033801
 AB033802 AB033803 AB033804 AB033805 AB033806 AB033807 AB033808
 AB033809 AB033810 AB033811 AB033812 AB033813 AB033814 AB033815
 AB036079 AB036080 AB036081 AB036082 AB036348 AB036349 BA000008
 VERSION AP002546.2 GI:10176692
 KEYWORDS
 SOURCE Chlamydomophila pneumoniae J138 (strain:J138) DNA.
 ORGANISM Bacteria: Chlamydiales: Chlamydiaceae: Chlamydomophila.
 1 (sites)
 REFERENCE Shiraishi, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,
 TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA
 JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)


```

1147 AAAGAGCTTTCTGAGTCAGACCCACTACCGAATCTCTATTACATCGG 1196
      :::::::::::::: ::::::::::::::
334  GluLysAspAlaIleuSerSerLeuGluAlaArgAsnGlyasp... 1 349
      :::::::::::::: ::::::::::::::
1197 AAACACCAATGCTAGATTGATACACCTCGAGCCTCTCAAGCAATACTA 1246
      :::::::::::::: ::::::::::::::
349  IeLeuPhePheAspProIleValGlnGluSerSerLeuGluSerPro 365
      :::::::::::::: ::::::::::::::
1247 TCTACTTCTATGATCTCTATVACA..... 1269
      :::::::::::::: ::::::::::::::
366  LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
      :::::::::::::: ::::::::::::::
1270 .....ACTAGATCACTGACACTCTCTCAGATGCTCTAACTTAATAGG 1313
      :::::::::::::: ::::::::::::::
382  AserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerS 399
      :::::::::::::: ::::::::::::::
1314 TCTTGACCTTCAGAGGAATCTCGCATATCAAGAACCATCGTATTTTCG 1363
      :::::::::::::: ::::::::::::::
399  eGluArgLeuSerGluGluGlyThr... ProAspAsnLeuThrSer 414
      :::::::::::::: ::::::::::::::
1364 GAGACAAAGCTTCGGAAAGCAGACAGCTCCGAAGCTGATTAATCTCAAA 1413
      :::::::::::::: ::::::::::::::
415  GlnLeuGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
      :::::::::::::: ::::::::::::::
1414 ACAATTCACCAACCTCTACCTCTCGGAGGCAACTCTCTTTAATC 1463
      :::::::::::::: ::::::::::::::
431  pArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuL 448
      :::::::::::::: ::::::::::::::
1464 AGGAGTCACCTAGTGTGTAAGCTTTTCGAATCTCCGGCTCTACCC 1513
      :::::::::::::: ::::::::::::::
448  euIleMetIuAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
      :::::::::::::: ::::::::::::::
1514 TCTCATGTCATCGAGGACACATTAAGAAACCGCTCATGTCGATCCTATC 1563
      :::::::::::::: ::::::::::::::
465  AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
      :::::::::::::: ::::::::::::::
1564 AATAATCTTGTCTCAATGATGATCTCTTAAAGAACCAAGCAAGCTAC 1613
      :::::::::::::: ::::::::::::::
481  rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
      :::::::::::::: ::::::::::::::
1614 GCTAAAGACACACAGCAAGCT...CAGACAGTCACCTTTATCTGATCCG 1660
      :::::::::::::: ::::::::::::::
498  .....GlyAspGlnAsnPheThrGluAsnValGluLeuLeu 509
      :::::::::::::: ::::::::::::::
1661 TCTCTCTTCTAGATCCTTCGAAATCTCTACGACAGATGTC..... 1701
      :::::::::::::: ::::::::::::::
510  SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLys 523
      :::::::::::::: ::::::::::::::
1702 .....TCTTGAATACCTCAAGCTTTTCTTCTCTCACTCTTACTGC 1745
      :::::::::::::: ::::::::::::::
523  sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
      :::::::::::::: ::::::::::::::
1746 TGAGACCCCGCGAATATTCACATCACAGAC.....TAGCTGCTGATC 1789
      :::::::::::::: ::::::::::::::
538  .....HisPheGlyThrGlnGlyAspThrThrPhe 547
      :::::::::::::: ::::::::::::::
1790 CCTAGAAAAAATCCTATCCATTCGGGATACCAAGGAAATGGGCATTA 1839
      :::::::::::::: ::::::::::::::
548  SerThrLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
      :::::::::::::: ::::::::::::::
1840 TCTTGCGCAAGAGTACTGCGCAATAATCCAAACAGCAGACTCTTACTGC 1889
      :::::::::::::: ::::::::::::::
562  rThrProLysAsnThrValProHisProGluLysGlnSerThrLeuValA 579
      :::::::::::::: ::::::::::::::
1890 GACAAAAACAGATACAAATCCGATCTCGACCTCGTGGAACTTAGTATG 1939
      :::::::::::::: ::::::::::::::
579  IaaenThrLeuThrAsnThrThrSerAspMetGlnAlaValGlnSerMet 595
      :::::::::::::: ::::::::::::::
1940 CTAAACCCCTTGGGGATCCTTCTGTGATGTGGCCTCATACAAACGCTT 1989
      :::::::::::::: ::::::::::::::
596  IleAsnThrIleAlaHisGlyAlaLysLeuPheGlyThrThrGlySer 612
      :::::::::::::: ::::::::::::::

```

```

1990 GTACCACTAAAGTACCCCAATCTCAAGAAACTGCGGCATCTGCTGTA 2039
      :::::::::::::: ::::::::::::::
612  rAlaValSerAsnLeuPheThrAlaHisAspSerSerGlyLysProIleA 629
      :::::::::::::: ::::::::::::::
2040 AGGATCTCGAATCTTTC.....CATAAAGATAGCAGCAAGATAATA 2083
      :::::::::::::: ::::::::::::::
629  sPasnThrHisIleArgSerLeuGlyThrLeuPheGlyIleSerThrHis 645
      :::::::::::::: ::::::::::::::
2084 AAGCTTTCCGCACATAGATGACAGTTATGTTGTAGAGCCCATCAACA 2133
      :::::::::::::: ::::::::::::::
646  SerLeuAspAsnHisSerPheCysLeuAlaIleAlaGlnLeuLeuGly 662
      :::::::::::::: ::::::::::::::
2134 TTAGCTCTGATTAATCTTATCACTGACCTGACGCTTCGCAATTAATCG 2183
      :::::::::::::: ::::::::::::::
662  sSerSerAspSerPheIleThrSerThrGluThrThrSerThrIleAla 679
      :::::::::::::: ::::::::::::::
2184 AGATAGACATCACTTTAAATAAATAAGACCTTCGCTATGAGCCTT 2233
      :::::::::::::: ::::::::::::::
679  hValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
      :::::::::::::: ::::::::::::::
2234 CTCTCCATCTCCAGCATCAGCAGCTTGTCTTCTCAAGCTTGTTCACGC 2283
      :::::::::::::: ::::::::::::::
691  .....IleSerAlaGlnAl 695
      :::::::::::::: ::::::::::::::
2284 TACCTTCTGATCTGAAAGTACAGCAGCTGCTCTTGTGATCTCAGAT 2333
      :::::::::::::: ::::::::::::::
695  acyThrAsnGluSerIleHisGluLeuLysThrLysThrArgSerPheS 712
      :::::::::::::: ::::::::::::::
2334 CAGCTATATCTATGTAATAAATACATATGAACCTTATTAACACCAAGC 2383
      :::::::::::::: ::::::::::::::
712  eLysGluGlyPheGlySerThrHisSerValAlaValSerGlyLysAl 728
      :::::::::::::: ::::::::::::::
2384 CAAG...CGACAGAGCTCGCTGTAAATGACGCTTGCCCTCGAAGCT 2430
      :::::::::::::: ::::::::::::::
729  CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
      :::::::::::::: ::::::::::::::
2431 GCGAGCTCCCTACACACACTGCTTAAACCAAGAGGCTTCTTCACCC 2480
      :::::::::::::: ::::::::::::::
744  rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGly 761
      :::::::::::::: ::::::::::::::
2481 GATTTTCTCTTATCAATGAAGTAGAACCTTCGATACACCAAGATAGCT 2530
      :::::::::::::: ::::::::::::::
761  heGluGlu...SerSerGlyLysIleArgSerPheSerAlaSerSerPhe 776
      :::::::::::::: ::::::::::::::
2531 TCAAGACAGTATATCTACTCTGCTGACGATCTTTCCATGACGGTGATTA 2580
      :::::::::::::: ::::::::::::::
777  ArgAsnIleSerLeuProIleGlyIleThrPheCysLysLysSerGln 793
      :::::::::::::: ::::::::::::::
2581 ATTAAGCTCTGTGCTGCTATGGAATTAACCTTGAGAGATCTCGAADA 2630
      :::::::::::::: ::::::::::::::
793  sThrArgThrThrThrPheLeuGlyAlaLysThrIleGlnAspLeuLysA 810
      :::::::::::::: ::::::::::::::
2631 CGAGCTGCTGCTTACCAAGCTACTGTCATCTAGCTGCGAGCTGATAC 2680
      :::::::::::::: ::::::::::::::
810  rGAspValGluSerGlyProValValLeuLeuLysAsnAlaValSerThr 826
      :::::::::::::: ::::::::::::::
2681 GTAAGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2730
      :::::::::::::: ::::::::::::::
827  AspAlaProMetAlaAsnLeuAspSerArgAla..... 837
      :::::::::::::: ::::::::::::::
2731 AAAACTACAGAAAGAAATCTCTCAAGCAAGCTGCTGCTGCTGCTGCT 2780
      :::::::::::::: ::::::::::::::
838  .....TyrMetPheArgLeuThrAsnGlnArgAlaLeuHisArgLeuS 852
      :::::::::::::: ::::::::::::::
2781 GATCTTTTATGCGCTTCT.....CCTAAATCTGCTG 2809
      :::::::::::::: ::::::::::::::
852  IThrLeuLeuAsnValSerCysValLeuArgGlyGlnSerHisSerThr 868
      :::::::::::::: ::::::::::::::
2810 AGTCACAAGTACCTATCTATGGAATTCGTGTGATCTTCAACGAGCTAC 2859
      :::::::::::::: ::::::::::::::
869  SerLeuAspLeuGlyThrThrThrArgPhe 878
      :::::::::::::: ::::::::::::::
2860 AATGACAGTCTTGAGGATAGTTCAGTTC 2889
      :::::::::::::: ::::::::::::::

```



```
RSVIFSSERLSEEEKTPDNLTSLQLOPIELKSGRIKXKRAVLASAPSLSDPOLLIIM
EACTSLKTSDDLKIALUSIPLHSLDTEKSYTHAPNISLSTOKITLSNSGDNFENIEL
LSKEONNIPLLTLSKEQSHLHPDGNLSHFQGGWTFESMKDSDESHLIAMWPERN
YVHPEROSTLVANTLANTYSDMQAVOSMINTLHGAGVLEGTWGSVSNLFLYAHDS
GKPIDMNHRSLSGLVLEGISTRHSDSHFCLAAGOLGKSDSEITSTETSYIATVOA
OLATPLMKISAQACYNESIHELKTKYRSFSGEGSMHSAVAGENCASIPIVNSG
LFSFSTLFSKLQGFSGTODPESSSGEIRFSFSSPNRNLSPMGITPEKKSOKTRANY
YFLGATYIDLRDVEGPPVLLKNVSWDAPMANLDSRAIMFRLTNORALHRLQTLIN
VSIVLRQSHSHSLDLGTYR1
3725..5500
/gene="Ct875"
3725..5500
/gene="Ct875"
/transl_start=1
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC68473.1"
/db_xref="GI:3329351"
/translation="MSTRGVGNGNRSKIPSNNGCNSRRSNTKNNKVEKVSILYS
SRSENRESPIYAVDVSMLESTPTSGETTRASKGVSRPQGLVRAADKRRAYVCCA
WSVSTRRSATRAEAGSSSRTRARGASSGREYSPSAARGLRLMFTDFMTRVLRQT
SPMAGVGNLDVNEARLMAATSECAHLEANKLAGDVAAREIAKRMQORVRLDIO
DKGAARKLNDPLGRTPNYOSKNRGEXTVNSMVEGPOVANTONVDGFLMSNL
SDVYLSREIQGLARATLESMPLENERFRRLOETCOARTELIEESGWTRESAS
RMEDEAGPSRAOQAFOSFVNECNSIEFSFGSGEHRYVCARVSRGLAAGAIRR
CFSCCKSGTHRYAPRDLSPEGASLAETLARFADMGIERGADGTYDPLVDDMRGV
PSIEGEGSDSIYEIMPIYEVMDMLETRSFAYQOQHYODPANSYDLPSPASVDLP
RSPPTPLPRYOLNDMDVEAGFREAVYASFAVMYVVTQOERLIPNSQVEGIL
RDMTNGSOTFRDLRMNRNREVDRE"
```

BASE COUNT 1664 a 1030 c 1382 g 1525 t
ORIGIN

alignment_scores:
Quality: 4509.00 Length: 878
Ratio: 5.147 Gaps: 0
Percent Similarity: 99.772 Percent Identity: 99.544

alignment_block:
US-09-677-752-4 x AE001361/rev ..

Align seq 1/1 to reverse of: AE001361 from: 1 to: 5601

```
1 MetArgProAspHisMetAspHisCysCysLeuCysAlaIleLeuSe 17
|||||
3451 ATGGACCTGATCATGAACTTCTGTGTCTATGCTGCTGCTATTTGTC 3402
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGlnThrAlaLeuL 34
|||||
3401 ATCCACACCGCTCCTTGGCCAGACCTCTAGGTGAAACGCCCTCC 3352
34 eutThrLysAsnProAsnHisValAlaCysThrPhePheGlnAspCysThr 50
|||||
3351 TCCTAAATAATCCTATCATCTCTCTCTACATTCTTTTTCAGACACTAC 3302
51 MetCysLeuPheProAlaLeuCysAlaHisAlaSerGlnAspAspPr 67
|||||
3301 ATGGAGACCTCTTCTGCTGCTGCTCATGATCACAAGATGATGATCC 3252
67 OleuTyrValLeuGlnAsnSerTyrCysTrpPheValSerLysLeuHisI 84
|||||
3251 TTTCTATCTACTTGGAAATTCCTACTGTTGGTCTGCTATCTAAACCTCA 3202
84 IetHisProLysGlnAlaLeuPheLysGlnLysGlnAspLeuSerIle 100
|||||
3201 TCACGACGCCCAAGACGCTCTTTTAAAGAAAAGAGATCTTTCATTT 3152
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGlnSerSe 117
|||||
3151 CAAATTTTTCGCTTCTCTTCCACAGATTGCTTCCCAAGAAAGCTC 3102
117 rProSerIleIleHisGlnLysAsnGlyClnLeuSerLeuArgAsnArg 134
|||||
```

```
3101 TCCTTCTATTTATTCATCAAAAGATGCTCACTTATCCTTGGCAATATG 3052
134 IySerMetSerPheCysArgAsnHisAlaGlnGlySerGlyValAlaIle 150
|||||
3051 GTAGCATAGTTTCTGTGCAATCATGCTGAGAGGCTCTGGAGAGACCTTC 3002
151 SerAlaAspAlaPheSerLeuGlnHisAsnTyrLeuPheThrAlaPheG 167
|||||
3001 TCTGGCGATGCTTCTTCTCTACACACACACTATCTTTCACAGCTTTTGA 2952
167 uGlnAsnSerSerLysGlnGlyValAlaIleGlnAlaGlnThrPheS 184
|||||
2951 AGAGAAATTCCTTCTAAAGAAATGCGGAGCCATTCAGCTCAACCTTCCT 2902
184 eLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAsp 200
|||||
2901 CTTTATCTAGAATGCTGCTGCTATTTCTTGCCGATATGTCGCGAT 2852
201 LeuAsnGlyValAlaIleCysCysSerAsnLeuIleCysSerGlyAsnVa 217
|||||
2851 TTAATGGCGGCGCTATTGCTGCTAGTATCTTATTCTTCAAGGAATGT 2802
217 IAsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyValAlaIleC 234
|||||
2801 AAACCTCTCTTTTTCACATGGAACCTCCGCCAGCAATGAGAGCGCTATT 2752
234 ySerLysIleSerAspLeuAsnThrSerGlnLysGlySerLeuSerLeuAla 250
|||||
2751 GTTGTATCAGCGATCTTAACACCTCAGAAAAGCGCTCTCTCTCTGCT 2702
251 CysAsnGlnGlnThrLeuPheAlaSerAsnSerAlaLysGlnLysGlyG 267
|||||
2701 TGTAAACCAAGAAAGCGTATTGTCAGCAATTTCTGTAAGAAAAGCGCG 2652
267 yAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyProValSerP 284
|||||
2651 GGCATATTATGCCAACACACATGCTATGCTTATTAACGGCTCTGTTCT 2602
284 helLeaAsnAsnSerAlaLysIleGlyValAlaIleAlaIleGlnSerGly 300
|||||
2601 TCATTAAACAACAGCGCTTAATAATAGGTGAGCTATCGCAACGATCCGGA 2552
301 GlySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGlnAsnAs 317
|||||
2551 GGGAGTCTCTCTATCCTTCTGCGAGTGAAAGATCTGTTCTTCAGAAATA 2502
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeuG 334
|||||
2501 CTCCCAACGACCTCCGACCAAGCTCTAGTAAGAAAGCCATCTACTTAG 2452
334 IuLysAspAlaIleLeuSerSerLeuGlnAlaArgAsnGlyAspIleLeu 350
|||||
2451 AGAAAGATGCGATCTTCTTCTCTTACAGCTCGCAACGAGATATCTT 2402
351 PhePheAspProIleValGlnGlnSerSerSerLysGlnSerProLeuPr 367
|||||
2401 TTCTTTGATCTTATGTACAAAGAAAGTAGCAGCAAGATGCGCTTCC 2352
367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSerP 384
|||||
2351 CTCCTCTTTCAGACCGACGCTGACTTCTCCACCCACGCCACCTCATCTC 2302
384 rOleuValIleGlnThrSerLysAsnArgSerValIlePheSerSerGlu 400
|||||
2301 CTTTAGTATTTCAGACGATGCAACGCTTCAAGTATTTCTTCAGAGCGAA 2252
401 ArgLeuSerGlnGlnGlnLysThrProAspAsnLeuThrSerGlnLeuG 417
|||||
2251 CGTCTTTTGAGAAAGAAAAGTCTGTATTAACCTCACTTCCCAACTACA 2202
417 nGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysAspArgAlay 434
|||||
2201 GCAGCCTATCGAAGTGAATCCGACGCTTATGTTTAAAGATCCGCTG 2152
```



```
373 ..GATTTCCTCTCTAGCACAAG...GTTACTACAGCTCAAGGACCT 416
130 uArgSnaSngIySerMetSerPheCysArgAsnHisAlaGluGlySerG 147
417 TTCCCTCAGCAGAGCCGTAATTTA.....GAAATATTTC 451
147 LyGIAlaIleSerAlaSpAlaPheSerLeuGln..... 158
452 GTAAACTGTAGTGTGCGGAATTTTTCATCGACAGATGCTGAGCTATCA 501
159 .....HisnTyrLeuPhe.....ThrAla.PheGluG 168
502 AAGGAGCGCTTTCTTCTTAACGCGACCTTCGAGATGCTCTTTTACGA 551
168 LuAsnSerSerLysGlySngIyAlaIleGlnAlaGlnThrPheSer 184
552 ACAACTCTTCATCACAAGAGGAGGAGCAATTCCTAAGACAGCCGCT 601
185 ..LeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAs 200
602 CCGCATAGCAAAATMACAGGTTATGTTAGATTCTTCTAATCATAGCGTC 651
200 PleuSngIyAlaIleCysCysSerAsnLeuIleCysSerGlyAsnV 217
652 TAGCTCAGAGCGCCTATCGATGATGCAAGCCACGCTCATATGCGACA 701
217 AlaSnpProLeuPhePheThrGlyAsnSerAla...ThrSngIyAla 232
702 ACAAAATTCTATATTGTAAGCGAATGCGAGCAAAACTACGCGGTCCG 751
233 IleCysCysIleSerAspLeuAsnThrSerGluLysGlySer.....Le 247
752 ATCTGC.....AACACAAAGCGAGTGTGATCTCTCGTCAAT 786
247 userLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysG 264
787 GATTAACTCTACAAATAACACTCTGATCTTCTTCAACGTAAGCAGAA 836
264 LuLysGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGly 280
837 CAAGCGGTGCGCCCATCCATCTCAAAACCTACCCCTTCTCTCGCAGCG 886
281 ProValSerPheIleAsnAsnSerAlaLysIle.....ClyG 293
887 TTTCAGAGCTTTCTAGAAATATGTCATCAGCAACTCTTAAGGGGG 936
293 yAlaIleAlaIleGlnSerGlyIySerLeuSerIleLeuAlaGluGln 310
937 TCGTATCAGATCGATGCTCAGAGAGCTCATCTTTCTCCAGACAG 986
310 LySerValLeuPheGlnAsnAsnSerGlnArgThr.....SerAspGln 324
987 GAAACTTACCTTTGTAAGAAATACCTTACAAACCGAAGTACCGCAT 1036
325 GlyLeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSe 341
1037 ACTCTTAACGTAATGCGATCAACTAGGAGCTAAGGGGAAATTCGCGGA 1086
341 rLeuGlu...AlaArgAsnGlyAspIleLeuPhePheAspProIleValG 357
1087 ATTACCGCTCTCTAAAAATCATACATTTTCTTCTATCATCCCATCAC 1136
357 InGluSerSerLysGlySerProLeuProSerSerLeuGlnAlaSer 373
1137 CAGAAGCACTCATCAGACGCTATTGAAAGATAATTAACGCTCTCCGCG 1186
374 ValThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSe 390
1187 GCTCTCAATCCATATCAAGGAAG..... 1210
390 rAlaAsnArgSerValIlePheSerSerGluArgLeuSerGluGln 406
..... 1211
1211 .....ATTCTATTTTCTGAGAAACCTTAACAGCAGATGAAC 1247
407 ..LysThrProAspAsnLeuThrSerGlnLeuGlnGlnProIleGluLeu 422
1248 TTAAAGTTGTCGCAATTTAAATTTCTTCTATCCAGCAGCCAGTCCCTA 1297
423 LysSerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSe 439
1298 TCCGAGGAAGTTTCTCTCAAAAGGAGTCACTTTAGAGAGCAGAG 1347
439 rLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerL 456
1348 CTTCCTCGAAGAGCGCGTCTCTCTCGCATGATTCAGAGAACGACAT 1397
456 euLysThrSerSer...AspLeuLysLeuAlaThrLeuSerIleProLeu 471
1398 TATCAACTACAGCTGGGAGTATTTCATACAGCAACCTAGATGATCANT 1447
472 HisSerLeuAspThrGluLysSerValThrIleHisAlaProAsnLeuSe 488
1448 GACTCCTTAAGCTTTAAGCAGCGCTCAGCTTAACGCAAAAGTGCTTC 1497
488 rIleGlnLysIlePheLeuSerAsn.....SerGlyAspG 500
1498 A...AAPAAGTGAATGATATGTCGCAAGCTCAACCTGATGATTTGAAG 1544
500 LuAsnPheTyGIuAsnValGluLeuLeuSerLysGluGln..... 513
1545 GGAACATTATGAAGT...CATATGTTCAAGCAAGACAGCTCTTCTCT 1591
514 .....AsnAsnIleProLeu 519
1592 CTATTAAATACAGCGTTGATGCTGATGATACAAAGTTGACATGAG 1641
519 uThrLeuSerLysGlnSerHisLeuHisLeuProAspGlyAsnLys 536
1642 CAGCCTT.....ATCCCTGCTCTCTGAGATGCTA 1673
536 erSerHisPheGlyTyrGlnGlyAspThrPheSerTrpLysAspSer 552
1674 ATTCAAGATACGATTCACAGACAAATGTAATGTAGGACTCGAGAT 1723
553 AspGluGlyHisSer.....LeuIleAlaAsnThrPheProLysAsnTy 567
1724 ACAGCTACAATACAAAGAGCCACGCGCACTTGACCAAAACGAGAT 1773
567 rValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrpA 584
1774 TGTCCCAAGCCCGAAGAAATGCGGTAGTATGCAATACCTATGCG 1823
584 snThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
1824 GAGTCTTACTGACATTCGCTCTGCAACAGCTTGAGAGATCGGCGCA 1873
601 HisGlyGlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLe 617
1874 ACTGGTATGGAACACAAACAGGTTTCTGCTTCTTCATACAGAAC... 1921
617 uPheTyrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHisA 634
1922 ....TTCCGTCATTAAGACTGAGATGAATAATGCAAGCGTCCGTCATA 1967
634 rGSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAsnHis 650
1968 CCTGTGGAGGCTACATCGATGGTGAAGGTGTCACACACTTAAGAAGCAG 2017
651 SerPheCysLeuAlaAlaGlyInLeuGluGlyLysSerSerAspSerPh 667
2018 CTATTACCTTTGCGTTCGCCATCTCTTGTGCTAGAGCAAAAGATTGTT 2067
667 eIleThrSerThrGluThr...ThrSerTyrIleAlaThrVal..... 680
2068 TATATCGCTCACACAAGACTAGAACSTAGACSTGAGAACCTTATCTTCAA 2117
```

895 CGTATCTATGCTAAGGAAATGTCACATTTTCTAATATACAGCAGCTGA 944
245
945 CTTTCCAAACAAATACACATCTCCACAACACTCTTACTGACACTTACA 994
245
995 CCTCCACCTACACACAGCAGTCACTCTTTGGTTAGATATGACAGCCC 1044
246 SerLeuS 248
1045 CATCTTCTGACTCTCCAGCTACCCCCCACAACAGGTGTTACCCCTGA 1094
248 erLeuAlaCyasnlnglnluthrLeuPheAlaSerAsnSerAlaLeuSgl 264
1095 CTATATCTGGAGAAACAGCCGTTCATTCCTAGAAACATTCCTCCCAA 1144
265 LysglYglYalAlaIleTyrAlaLysHisMeValLeuAgtYrAsnGlyPr 281
1145 CAACGACGACCCCTCTATGCCAAAAGATCTCTATAGATTCATAATAAATC 1194
281 ovalSerPheIleAsnAsnSerAlaLysIleGlyYglYalAlaIleG 298
1195 TACATATATTTCTTGGAAATACAGCTGGAAGAGCGCTATGCTATGTC 1244
298 InSerClYglYSerLeuSerIleLeuAlaGlyClYglYSerValLeuPhe 314
1245 CCGAATTCGCGAGAGCTCTCTATGCCCAATACAGGTGATATCTCTTTT 1294
315 GluAsnAsnSerGlnAgtThrSerAspGlnGlyLeuValAgtAsnAlaI 331
1295 AACAGAACCTCAGCATCTACTGCGACACTT...ACTGCCAATATGAT 1341
331 eYrLeuGluYasAspAlaIleLeuSerSerLeuGluAlaAgtAsnGly 347
1342 TCACCTTCGGAAGATGCGAAGTTGCACTCTAGGAAATACGCAAGGT 1391
348 ..AspIleLeuPheAspProIleValGlnGlnSerSerSerYsglu 363
1392 ATACCCCTATACCTTATGATCCGATTTACATGATGATTTATCTGCTGA 1441
364 SerProLeuProSerSerLeuGlnAlaSerValThrSerProThrProAl 380
1442 TCC..... 1444
380 aThrAlaSerProLeuValIleGlnThrSerAlaAsnAgt..... 393
1445 ..GCAGCCGCTACTGTGTCGTCAATCCCAAGCCAGTGCAGATGTCGCT 1493
394SerValIlePheSerSerGluAgtLeu...SerGluGluGlu 406
1494 ATTCAAGGAGCTATGCTTTTCAGAGAAACCTCAGCTACCTACCAAGCA 1543
407 LysThrProAspAsnLeuThrSerGlnLeuGlnProIleGlnLeuLys 423
1544 GCAACCCCTGCAAAATGCTACTACATTAACCAAAAGCTAGAACTTGA 1593
423 sSerGlyAgtLeuValLeuLysAspAgtAlaValLeuSerAlaProSerL 440
1594 AGCGGCTACTCTGCTTTAAGAACGGTCTACTTAATGTTCAATACT 1643
440 euSerGlnAspProGlnAlaLeuLeuIleMeGlnAlaGlyThrSerLeu 456
1644 TCACGCAAGATGAAAGTCGCTGTCATCATGATGATGACAGGACCAATTA 1693
457 LysThrSerSerAsp.....LeuLysLeuAlaTh 466
1694 CCACTACCAAAATGAGCTAATAACTGACCGCTGCTACCTTAACCA 1743
466 rLeuSerIleProLeuHisSerLeuAspThrGluYsgSerValThrIleH 483
1744 GCTTGTATCATCTGGATTTCTTGGATGGCACTAAAGCGGCTGTGTTA 1793

483 lAlaProAsn.....LeuSerIleGlnLysIlePheLeuSer 495
1794 ATGTGACAGATACCAATGGACCTCTACATATATCCGGACCTTTAGACTT 1843
496 AsnSerGlyAspGluAsnPheTyrGlnAsnValGluLeuLeuSerLysG 512
1844 GTGAAAACCTCTCAAGATTGCTGTGACACACCGGAGATGTTAATAAGA 1893
512 uGlnAsnAsnIleProLeuLeuThrLeuSerLysGluGlnSerHisLeuH 529
1894 TTTCAGCAGATCCGATTTTGAATCTAAAGCACTTCAATACTGTAA 1943
529 lAlaProAsp.....GlyAsnLeuSerSerHisPhe 539
1944 CCACATACGAGATTCAGTCTCGCAAAACGGCTATACAGCATTCGCCCAT 1993
540 GlyTyrGlnGlyAspTrpThrPheSerTrpLysAspSerAgtGluGlyH 556
1994 GGTATACAGAGACTTGGAGTTTACATACATACAGCAACAC.....CA 2037
556 sSerLeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGlu 573
2038 TACGGTCACAGAAATTTGAAAAAACCGGTTATCTCTTCTATCCGAGC 2087
573 rglInSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMet 589
2088 GTCCTGCTGCCCTCATCTTCATATAGCCATAGCGCAACCATCATGATTTA 2137
590 GlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
2138 CGAGCTGTAACTCAAGCG.....TCAGCAGCTGTAGCGGAAGATGTC 2181
603 ..GlyAlaTyrLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPhe 619
2182 TGGGAGACACTGAGACATCACA.....GGAATTCAAAATTTCTTCC 2222
619 yAla...HisAspSerSerGlyLysProIleAspAsnTrpHisHisArg 634
2223 ATCGAATCAATACCGGTATGCACGC.....ACCTACCGCAATG 2263
635 SerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHis 651
2264 GGTGGAGGCTACTTCATCAATACCTACACACGACATCACTCCAGATGTC 2313
651 rPheCysLeuAlaAlaGlnLeuLeuGlyLysSerSerAspSerPhe 668
2314 GTTAAGTCTAGGTTTGGACACAGCTGTTACAAAATCTAAGATTACCTCG 2363
668 lThrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeu 684
2364 TAGTCACGATCATCTTAACGTTTATTCGTCACAGTATACCTTAACATC 2413
685 AlathrSerLeuMetLysIleSer.....AlaG 694
2414 ACCAAGTCTCTGTTGATGATCGAGATTTCTTCAGAGACACTTCG 2463
694 nAlaCysTyrAsnGlnSerIleHisGluLeuLysThrLysTyrAgtSerP 711
2464 ACTTACCTATACCCCTACCAATGAGAAAGTAAGCTTCATATACAAAT 2513
711 heSerLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGlu 727
2514 TCCCTAA...GGCGCGTCTCTTGGAGTAACAATGCTGGTTAGAGAA 2560
728 ValCysAlaSerIleProIleValSerAsnGlySerGlyLeu...Phe 743
2561 CTCGAAGGAACTTCCATCACTCTCTCTTCCATCTTAACCTCA 2610
743 rSerPheSerIlePheSerLysLeuGlnGlyPheSerGlyTrpGlnAsp 760
2611 GCAGTATCTCCCTTTGTAAGCTGAAGTTCCTTACCGACATCTAGGG 2660

[illegible]

493 pheIeuSerAsnSerGlyAspClnuSnPheTyrgLnSnValGluLeuLeu 509
1606 TTAGCACTTGTGTAATGAGGAGGAGCTTTGTATGAAACCATACCTTGCA 1655
509 userLySGluGlnAsnAsnIleProLeuLeuThr..... 520
1656 AGACGCTCAAGCTTATATACCTTTGTAACCTTATCTCCAGCAACCGGAGCGCA 1705
521LeuSerLySGluInSerHisLeuHisLeuProAspGlyAsnLeu 535
1706 CTATATATTAATCTCAAGATCGCTTCTCGAAGAGCTTTTGAAGTAGTACCTCTTCT 1755
536 SerSerHisPheGlyTyrgLnGlyAspThrPheSerThrPheSerThrLyAspSer 552
1756 AACACCAATATTAGCTATCCAGAGACATTGGAAATGGCAAGTCATCCAGC 1805
552 rAspGluGlyHisSerLeuIleAlaSn.....TrpThrProLyAsnT 567
1806 AACGGCAACTCAACCGAGCCAGCCAAATTTAGAAATGGCTGCGGACAGAT 1855
567 yValProHisProClnuArgInserThrLeuValAlaAsnThrLeuThr 583
1856 ACCCTTCGCAATCCGACAGAGGCAAGGATTTTATGTTCCAAATAGGCTGTGG 1905
584 AsnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAl 600
1906 GGTCTCTTTGTGTGATCAGACGCTGCTATCCAGAAATTCATGTAATATAGAG 1955
600 ahIScLyGlyAlaTyrgLeuPheGlyThrTrpGlySerAlaValSerAsnL 617
1956 CCAAACTCTTATCTCAGAAACGGGAGCTGGGAGAGCTGCAATGCTTAAT 2004
617 euPheTyrgLnHisAspSerSerGlyLySProIleAspAsnThrPheHis 633
2005TTCCCTACATGAGAAATTAATTAATGAG.....CACGGCTATCGCAT 2046
634 ArgSerIeuGlyTyrgLeuPheGlyLeSerThrHisSerLeuAspAsnI 650
2047 ACGGCTGTGGTATCTCTTGTGGAATTTGGCACTAGCTATGCTTTTCTGATGC 2096
650 sSerPheCyLeuAlaIleGlyGlnLeuLeuGlyLySerSerAspSer 667
2097 TACGATTAATGCGCTTTTGGCCACCTCTCAGTAGAGATTAAGACATACG 2146
667 heIeuHisSerThrGlnuThrTrpSerTyrgIleAlaThrVal..... 680
2147 TAGTATCCAAAATCATGAGACTACTACTCAGGGGTGATTTCTTGAG 2196
681 61 681
2197 GATACCTTGAAGTTTAGAAGTCCACAGGATTTCTATCTGATAGCTCTC 2246
681 nAlaGlnLeuAlaThrSerLeuMetLyIleSerAlaGlnAlaCyTyra 698
2247 ACAAGCTTGCTGTACCAAGTCTGCTACATATGATATGCAAGTGGCTTACA 2296
698 snGluSerIleHisGluLeuLySThrTrpTyrgSerPheSerLySglu 714
2297 GCCATATGAATATGATGATGTAAGAAACCAATATACGACATAT...CCGAA 2343
715 GlyPheGlySerThrPheHisSerValAlaValSerGlyGluValCySAlaSe 731
2344 GCCACAGGATCTTGGCAATATATGTTTGTGCTTGAGTTTGGAGCGAC 2393
731 rIleProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIleP 748
2394 TACATACCTACTACCCCAACACAGTACTTTTATTTATTTATACATACCTCCGT 2443
748 heSerLySLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSer 764
2444 TTCTCGGTGGTGAAGTCCACTTATGCTTCCACCAAGAAAGCTTCAAAAGAGCA 2493
765 SerGlyGlnIleAlaSerPheSerAlaSerPheAlaGlnIleSerLe 781

```
578 CCGCCGACAAAGGGGACCATCTAAGCTTCTGCTGACTCTGTATTTCT 627
187 ArgAsnValSerProIleSerPheAlaAlaArgAlaAspLeuAsnI 203
      :::::::::::::::::::::
628 GGAATGCGAGGAAATCTGAGCTTGGAAACAAAGTGCAGCAACATCAAG 677
203 yGIyAlaIleCysCys...SerAsnLeuIleCysSerGIyAsnValAsp 219
      :::::::::::::::::::::
678 AGCCCGAGCTCTGCTGAGAGGAGACCTTGTGATCTCCATTAACCAAAAT 726
219 rOleuPhePheThrGIy...AsnSerAlaThrAsnGIyAlaIleCys 234
      :::::::::::::::::::::
727 ..ATCTTTTCGATGCTGCACAAAGCACTACAAATGCGGAGCTATTGAT 774
235 CysIleSerAspLeuAsnThrSerGIyLysGIySerLeuSerLeuAla 251
      :::::::::::::::::::::
775 TCT...AACAAAGCAGGGGCGAACCAGACCTATCTGACTCTTCAGG 821
251 sAsnGIyLysThrLeuPheAlaSerAsnSerAlaLysGIyLysGIyAla 268
      :::::::::::::::::::::
822 AATGAGAGCCGTCATTTCTGATATACACACAGCAAGAAATAGTAGAGTG 871
268 IaIleTyAlaLysHisMetValLeuArgTy...AsnGIyProValSer 283
      :::::::::::::::::::::
872 CGATTATACCAAAAATTCGTGTATCTCTCAGAGCAGCAGAGCTCTTA 921
284 PheIleAsnSerAla...LysIleGIyAlaIleAla 296
      :::::::::::::::::::::
922 TTTTCTAACACAAAGCTGCGAATCTCTACTCTTAAGAGGGGCAATTGC 971
296 aIleGIySerGIySerIleSerIleLeuAlaGIyLysGIySerVal 313
      :::::::::::::::::::::
972 GATTCTAGATTCTGAGAGATTAGCATTTCTGACATCTCGGCATATCA 1021
313 eUpHeGIyAsnAsnSerGIyAlaArgTy...SerAspGIyLysLeuAla 327
      :::::::::::::::::::::
1022 TTTTCCAGGCGCAATCTACGACACTACAGAGAGTCGCGAGTGAGAC 1071
328 ArgAsnAlaIleTyLeuGIyLysAspAlaIleLeuSerSerLeuGIyAla 344
      :::::::::::::::::::::
1072 ACAAATGCTATGATCTTGATCGAATGCAAAATTTTAAATCTCGGAGC 1121
344 aArgAsnGIyAsp...IleLeuPhePheAspProIleValGIyLysIle 360
      :::::::::::::::::::::
1122 GACTCGGGGAATAAGTTATTTCTATGATCTCTATC... 1158
360 eSerLysGIySerProLeuProSerSerLeuGIyAlaSerValThrSer 376
      :::::::::::::::::::::
1158 ..... 1158
377 ProThrProAlaThrAlaSerProLeuValIleGIyThrSerAlaAsnAr 393
      :::::::::::::::::::::
1159 .....ACGAGCTCAGGACCTCTGATAGCTCTCTTTGAAATTA 1196
393 gSer.....ValIlePheSers 399
      :::::::::::::::::::::
1197 AGCTACGCGAGATCTGGAATACCTATGAGGCTACATCTTTCTCTCG 1246
399 eArgIaArgLeuSerGIyLysLys...LysThrProAspAsnLeuThrSer 414
      :::::::::::::::::::::
1247 GAGACAACACTCTCAGAGACGAACTTAAGAAACCTGACAATCTCAAGCT 1296
415 GIyLeuGIyLysProIleGIyLeuLysSerGIyArgLeuValLeuLysAs 431
      :::::::::::::::::::::
1297 ACATTACACAGGCTGTAGAGCTTCTGAGGTGCTTGTAGTATTGAAGA 1346
431 PaArgIaValLeuSerAlaProSerLeuSerGIyAspProGIyAlaLeuL 448
      :::::::::::::::::::::
1347 TGAATGACTAGTGTGCAAAATACATATACGAGAGCTCGAGGATCGAAG 1396
448 eUIleMeGIyAlaGIyThrSerLeuLysThrSerSerAsp...LeuLys 463
      :::::::::::::::::::::
1397 TCGTATATGATGAGGAGACTCTTTGAGGCGCAAGCGCTGACGGGCTACT 1446
464 LeuAlaThrLeuSerIleProLeuHisSerLeuAsp...ThrGIyLysSe 479
      :::::::::::::::::::::
1447 CTCATATGGCCCTAGCCATTATATAGATTCCTTAATGAGCAAAATAAGC 1496
479 rVal.....ThrIleHisAlaLeuAsnLeuSerIleGIyLysIleP 493
      :::::::::::::::::::::
1497 TATCATTAAGGCGACCGCCAGCAAGTACAGATGCTGCTTATACGGCCCA 1546
493 hELeuSerAsnSerGIyAspGIyAsnPheTyGIyLysAsnValGIyLeuLeu 509
      :::::::::::::::::::::
1547 TCATGCTTGTATAGTGTCTCGGGGAAGCTATTATGAGATCATATCTC... 1593
510 SerLysGIyLysAsnAsnIleProLeuLeuThrLeuSerLysGIyLys 524
      :::::::::::::::::::::
1594 ...AGTCAACACAGAGTCTTCTCTTAATAGAGCTTCTGCAACAAGAAC 1640
525 ..GIySerHisLeuHisLeuProAspLys.....AsnLeuSerHis 538
      :::::::::::::::::::::
1641 GATGACTACTACAGATATCCCGGATACCCCATTTCTAAATACTACGAATC 1690
538 IaSPheGIyTyGIyLysAspTrpThrPheSerTrpLysAspSerAspGIy 554
      :::::::::::::::::::::
1691 ACATAGGATATCAGAGAACTGCAATATATCTTTGGGTGACAGATGCAACT 1740
555 GIyHisSerLeuIleAla...AsnTrpThrProLysAsnTyValPr 569
      :::::::::::::::::::::
1741 GCAAAAACAAAAAATGCTACTTACTTACTGACTTAAACAGATACAAAGCC 1790
569 oHisProGIyLysArgIleSerThrLeuValAlaAsnThrLeuTrpAsnThr 586
      :::::::::::::::::::::
1791 GAATCCAGAGACTCAGAGCACTTTGGTTCCTTAATACCTCTGGGTTCTT 1840
586 YrSerAspMetGIyAlaValGIySerMetIleAsnThrIleAlaHisGIy 602
      :::::::::::::::::::::
1841 TTGTGATGTCCGCTCCATTCACAGCCCTCATGACCGGAGCAGCAAGTTG 1890
603 GIyAlaTyLeuPheGIyThrTrpGIySerAlaValSerAsnLeuPheTy 619
      :::::::::::::::::::::
1891 TTAATCTTGCTACAAATTTGTGGTATCAGGAATCGCGGAC...TTTTT 1937
619 rAlaHisAspSerSerGIyLysProIleAspAsnTrpHisIleArgSerL 636
      :::::::::::::::::::::
1938 GCATGAAGATCAAGAAAGGAACCA...CGTATGATATCTCTATTCTACGC 1984
636 eUGIyTyLeuPheGIyLysSerThrHisSerLeuAspAspHisSerPhe 652
      :::::::::::::::::::::
1985 CGGCTTATGATATAGAGAGGATTTCTTACCGCTTGAATAATTTCTTT 2034
653 CysLeuAlaIleAlaGIyLysLeuLeuGIyLysSerSerAspSerPheIleTh 669
      :::::::::::::::::::::
2035 AATTTGCTTTTGTGACGCTTTTGGCTACGACAAGGACCATCTGTGGC 2084
669 rSerThrGIyLysThrSerTrpIleAlaThrValGIyAlaGIyLys 683
      :::::::::::::::::::::
2085 TTAGAACCATACCCATGTATATGACAGGGCAATGATACCGACACCTCG 2134
684 ..LeuAlaThrSerLeuMetLysIle..... 691
      :::::::::::::::::::::
2135 GAGACTCTAAGACCTCGCTAGATTTTGTGCAAAATTCGACTCCCTA 2184
692 .....SerAlaGIyAlaCysTyArgAsnGIyLysIleHisGIyLeu 704
      :::::::::::::::::::::
2185 CTTTGTGCTTCAATGCTCGGTTTGTGCTTATGCGCATACGCAATAATACAT 2234
704 uLysThrLysTyArgSerPheSerLysGIyLysPheGIyLysTrpHis 721
      :::::::::::::::::::::
2235 GACCCAAAGATACACTGGCTATCTCTCTGTTAAG...GGAAGCTGGGAA 2281
721 eValAlaValSerGIyLysValCysAlaSerIleProIleValSerAsn 737
      :::::::::::::::::::::
2282 ATGATGCCCTTCGTTATGAATGTGAGAGAGATATCCGTTAGTTCTTCA 2331
```

```
487 GGGCCCATCATCGCTAAATCACAACCCCTCAAGAAAGCTGCATTAAAT 536
124 yAsnGlyGluLeuSerLeuArgAsnAsnGlySerMetSerPheGlyAsp 140
537 GTGGAGGGGATCTTACATTTCATACAAATGCAATATTTTATTAAACAA 586
141 AsnHisAlaGluGlySerGlyAlaIleSerAlaAspAlaPheSerLe 157
587 GATTACTGTGAGAAATAGCGGAGCCATTCTCCAGAAATCTTCTTT 636
157 uClnHisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerGly 173
637 GAAAAACACGACGGCATCTTCTTTGAAAGCCAAATTAATCCACCAG 686
174 ..... AsnGlyAlaIleGlnAlaGln... ThrPheSerLeuSer 186
687 CAGCAAAAAGGTGGCTATTTGCTCTACTGTAGTACATATATACAA 736
187 ArgAsnValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnG 203
737 AATATACGGCTCTACCTCTCTCCAGAACATATGCTGAAAGCTGCAG 786
203 yGlyAlaIle... CysCysSerAsnLeuIleCysSerGlyAsnValAsp 219
787 TGCAGCTTAAATAGCACAGCAAACTCTACAAATTCACAGCAATACCT 835
219 rOlLeuPhePheThrGlyAsnSerAlaThr..... AsnGlyGly 231
836 ..CTGTATTTCTGAAATAGCTGACAGCAGCCAGCAAGAAATCCAGCA 883
232 AlaIleCysCysIleSerAspLeuAsnThrSerGluLysGlySerLeu 248
884 GCTCT..... TCTGGAGATGCCGATGTTAC 909
248 rLeuAlaGlyAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
910 CATATCTGGAGATCAGACTCTACTTCTCAGGAACCACTGACTAGCTA 959
265 ySGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGly... 280
960 ATGGCGGAGCCATTATCTCTAAGAAGCTTACACTGCTCCGGGGGGGG 1009
281 ..... ProValSerPheIleAsnAsn..... SerAl 289
1010 GGGGGGAATCCC..... TTTCTAACATATAGTCCAAAGTACCACCTGC 1053
289 aLysIleGlyAlaIleAlaIleGlnSerGlyGlySerLeuSerIleL 306
1054 AGTAATGCTGAGCCATTTCTATCTAGCTGACGCTGACAGTGTACTCT 1103
306 euAlaGlyGluGlySerValLeuPheGlnAsnAsnSer... GlnArgThr 321
1104 TCAGCGAAGCAGGGGACCATTAACCTTAATGGGAATCCCATTTGCAACT 1153
322 SerAspGlnGlyLeuValArgAsnAlaIleTyrLeuGluLysAspAla 11 338
1154 ACACACACAACTACAAAAGAAATCTATTACATAGATCTACTGGCAA 1203
338 eleuSerSerLeuAlaArgAsnGly... AspIleLeuPheAsp 354
1204 AGATCAGCAATTAAGTCAATATCTGGCAGATACATCTTTTCTACAGAT 1253
354 rOlLeuAlaGlnGlySerSerLeuSerLysGlySerProLeuProSerSerLeu 370
1254 CGATTACTGCCAATACGGCTCCGATCTCTACAGATCTTTTAAATCTCAAT 1303
371 GlnAlaSerValThrSerProThrProAlaThrAlaSerProLeuVal 11 387
1304 AAGCTGATGAGGTAAAT..... 1321
387 ecLlnThrSerAlaAsnArgSerValIlePheSerSerGluArgLeuSerG 404
1322 .AGTACAGATTATAGTGGGCGATTGTTTTTCTGCTCAAAAAGCTCTCTG 1370
404 LuGluGlu... LysThrProAspAsnLeuThrSerGlnLeuGlnIlePro 419
1371 AAGATGAAGCAAAAGCTTCGACCAACCTCCTACCTTCTACCTGAGAGCT 1420
420 lIleGluLeuLysSerGlyArgLeuValLeuLysAspArgAlaValLeuSe 436
1421 GTACTCTTAAGTGCAGGAATTTAGTACTTAACCTGCTGTCTACTCTGA 1470
436 rAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMetGluAla 453
1471 TACGAAGACCTTTACTACTAGACGGCGGCTTCTCTGTTTATGATGGCG 1520
453 lYlnrSerLeuLysThrSerSer... AspLeuLysLeuAlaIleThrLeuSer 468
1521 GCACAACTTAAAGCAAGTACAGAGAGGAGTACACTTTAACAGTCTTCC 1570
469 lIleProLeuHisSerLeuAspThrCulysSerValThrIleHisAlaIpr 485
1571 ATTCTGTAGACTCTTTAGCGAGGGTAAAGTTGTAATGTGCTGCTTC 1620
485 O..... AsnLeuSerIleGlnLysIlePheLeuSerA 496
1621 TGCAAGCAAGTAAAAATAGGCCCTTAGTGTCCTTCTTTTTCGATTA 1670
496 snSerGlyAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGlu 512
1671 ACCAAGG..... AATGCTTATGAAATCAGCAGC... TAGGAAAAACT 1711
513 GlnAsn..... AsnIleProLeuLeuThrLeuSerLysGlu 525
1712 CAAAGCTTTTATTTGTCACCTCTCTGCTGCTGCTGCTGCTGCTGCTG 1761
525 nSerHisLeuHisLeuProAspGlyAsnLeuSerSerHisPheGlyTyrG 542
1762 AGATGTTCCACCGGCTTCTACAGTACGCAATCTCAGCATAGTGGATC 1811
542 lngLysPheThrPheSerThrPlyAsp..... SerAspGlu 554
1812 AAGTACTTGGGAATGACCTTGGTGTATACCGCAAGCAGCCCAAG 1861
555 GlyHisSerLeuIleAlaAsnTyrThrProLysAsnTyrValProHisP 571
1862 ACTAAGACAGGACATTAAGCTTGAGCAATACAGCTTCCGGAATCC 1911
571 oGluArgGlnSerThrLeuValAlaAsnThrLeuTyrAsnThrTyrSerA 588
1912 TGACGCTCAAGACCTTATAGTCTCAATAGCCTTGGGAGATCTTTTCA 1961
588 spMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGlyAla 604
1962 ACATCCAAAGCATTCAAAGGTGTCTATAGAGAGAGTGTGACTCTTGT 2011
605 TyrLeuPheGlyThrTyrGlySerAlaValSerAsnLeuPheTyrAla 621
2012 TCAATGAGGCTTCTGGCGCGGAGTCCGCAAT... TTCCTTAAGTAA 2058
621 sAspSerSerGlyLysProIleAspAsnTyrPHisIleArgSerLeuLyt 638
2059 AGATAAGAAAGGCGGAAAA... CCGCAATACCGTCATAAATCTGCTGAT 2105
638 yrLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPheCysLeu 654
2106 ATGCTATCGAGGTGAGCGCAAACTGTCTGAAACTTAATAGCTTT 2155
655 AlaAlaGlyLlnLeuLeuGlyLysSerSerAspSerPheIleThrSerTh 671
2156 GCCTTTGGCAACTCTTGTGAGGATAAAGATTTCTTACGCTTAAATA 2205
671 rGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaIleSerL 688
2206 TCATACTGATACATATGACAGAGCCTTCTATATCCACACATTCAGAAAT 2255
```


1005 GAAATATATCTCATACACCACTCTTAAGGGGGGCTCATAGATCAT 1054
:::|||||
299 SerGlySerLeuSerIleuAlaGlyLeuSerValLeuPheG1 315
:|||||
1055 GCGTCAGAGAGACTCAGCTCTTCTCCACAGACAGGAACATTACCTTTGT 1104
:::|||||
315 nasnasSerGlnArgThr.....SerAspGlnCylLeuValArgAsnA 330
:|||||
1105 AAGAAATACCTCTTACCAACCGGAAGTCGAGTCTCTTAAGCTATAG 1154
:|||||
330 IalIeYLeuGluLysAspAlaIleLeuSerSerLeuGlu...AlaArg 345
|||||
1155 CGATCACAATGAGCAAGTACCGGAATTTCACGCAATTACGGCTCTCAAA 1204
:|||||
346 AsnGlyAspIleLeuPhePheAspProIleValGlnIleuSerSerLeY 362
|||||
1205 AATCATACAAATTTTCTTCTATATGATCCATTCACACTTCAGAACGACTCATC 1254
:|||||
362 sGluSerProLeuProSerSerLeuGlnAlaSerValIThrSerProThrP 379
:|||||
1255 AGACGATTGAGAGATTAATTAACGGCTCGGGAGCTCTCAATCCATCATC 1304
:|||||
379 roAlaIhrAlaSerProLeuValIleGlnIhrSerAlaAsnArgSerVal 395
:|||||
1305 AAGAGACG.....AArg 1315
:|||||
396 IlePheSerSerGluArgLeuSerGluGluGlu...LysThrProAspAs 411
:|||||
1316 CTATTTTCTCGAGAAACCTTACACACACATGCAACTTAAAGTTCGTCACAA 1365
:|||||
411 nleuThrSerGlnLeuGlnGlnProIleGluLeuLysSerGlyArgLeuV 428
|||||
1366 TTTAATATCTTATTCACCCAGCCAGTCCTCCCTTCCGAGGAAGTAT 1415
:|||||
428 alLeuLysAspArgAlaValIleuSerAlaProSerLeuSerGlnAspPro 444
:|||||
1416 TCCTACAAAGGAGACGACTTGAAGACACACCTCTCTCAAGAGGCC 1465
:|||||
445 GlnAlaLeuLeuIleMetLysAlaGlyThrSerLeuLysThrSerSer... 460
:|||||
1466 GCTTCTCTCTCGCCATGGATTCAGAGAACGACATTATACACACACGTCG 1515
:|||||
461 AspleuLysLeuAlaIhrLeuSerIleProLeuIhrSerLeuAspIhrG 477
:|||||
1516 GAGTATTACAATCAGCAACCTAGATGAATGTGACTCTTGAATCAAG 1565
:|||||
477 IuLysSerValIhrIleHisAlaProAsnLeuSerIleGlnLysIlePhe 493
:|||||
1566 AGCAGCCGCTCACGCCCTTAACAGCAAAAGGCTCA...AAATAAGTGATG 1612
:|||||
494 LeuSerAsn.....SerGlyAspGluAsnPheTyrGluAs 505
:|||||
1613 GTATCTGGGAACCTCAACCTGATTCATATTGAAGGGAACATTATTATGAAG 1666
:|||||
505 nvaIgluLeuLeuSerLysGluGln..... 513
:|||||
1663 T...CATATGTCAGGCATGACACCTCTTCTCTATTAATAATCAAGG 1705
:|||||
514AsnAsnIleProLeuLeuIhrLeuSerLysGlu 524
:|||||
1710 TTGATGCTGATTTGTGATACAACTTGATCAGCAGCGCTT..... 1750
:|||||
525 GlnSerIhrLeuIhrLeuProAspGluAsnLeuSerSerHisPheGlyTY 541
:|||||
1751ATGCCCTGCTCTGAGGATTCCTTAATTCGAATTAGCATTT 1799
:|||||
541 rGlnGlyAspTrpThrPheSerTrpLysAspSerAspGluGlnHisSer... 557
:|||||
1792 CCAAGGACATGGAATGTTAATATGGCACTACAGCTACAAATAATACA 1841
:|||||
558LeuIleAlaAsnTrpThrProLysAsnTyrValProHisProGlu 572
:|||||

```
1111 .....1111 1111 .....1111 1111 1111
1260 CCGCAAGTACGAGGACTTCTGCAAGTCAAGTAACGCTATCTATTTC 1309
352 PheAspProIleValGlnGlnSerSerSerIleGluSerProLeuProSe 368
1310 TATGATCCCAT..... 1321
368 rSerLeuGlnIleSerValThrSerProThrProAlaThrAlaSerProL 385
1322 .....ACCACCAAGATACGGAGCAAGCAAGCAATCACTTAC 1355
385 euValIleGlnThrSerAlaAsnArg.....SerValIlePhe 397
1356 GTATCAATAGAGCTCAGTCAAAATCAAAAGCTTCGAGATCTATAGTATTT 1405
398 SerSerGluAlaGluSerSerGluGluGluLysThrProAspAsnLeuThrSe 414
1406 TCTGGAGAGAGATTGTCACACACAAAGCTTACCTGAAATCTTACCTTC 1455
414 rGlnLeuGlnGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysA 431
1456 GAGGATCAACACGCTGTCACTTATGTAAGGGAGGAGCTTACTTAAAC 1505
431 sPArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
1506 AUGGAGTCACTTGATGCACACAGAGATTCTCGACAGACCAAGATCCACG 1555
448 LeuIleMetGluAlaGlyThrSerLeuLysThrSerSer...AspLeuLys 463
1556 CTTCCTTTGGATCTGGGACCTCAATTAAGCTTCTACAGAAAGATATTCCT 1605
463 sLeuAlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerV 480
1606 CATCAAAATTTATCTATTAATGCCGATACCATTTACGGAAGAAATCTCTA 1655
480 aThrIleHisAla.....ProAsnLeuSerIleGlnLysIle 492
1656 TCAATATTGTAGCTTCACACAGGATAGAACATTAACCTTACACAGCAAGC 1705
493 PheLeuSerAsnSerGlyAspGluAsnPhetyrGluAsnValGluLeuLe 509
1706 TTACGACTTGTAAATGACATGACAGCTTTCATGAGAACCAATACCTTCGA 1755
509 userLysGluGlnAsnAsnIleProLeuLeuThr..... 520
1756 AGACTCTCAAGACTATAGCTTTGTAAAGTTATCTCCAGAGCGGAGGGA 1805
521 .....LeuSerLysGluGlnSerHisIleuHisLeuProAspGlyAsnLeu 535
1806 CTATATATTACTCAAGATGCTCTCAGAACCCCTTGAAGTACCTCTCTCT 1855
536 SerSerHisPheGlyTyrGlnGlnLysPThrPheSerTrpLysAspSe 552
1856 AGACCAATATAGCTATCAAGACATTCGAATTCGCAAGTCAATCCCAAG 1905
552 rAspGluGlnHisSerLeuIleAlaAsn.....TrpThrProLysAsnTr 567
1906 AACGGGAACTCAACGACGACGACAAATTTAAATGGGTGGCGACAGGAT 1955
567 TyrValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrp 583
1956 ACCTTCCGATCCGACGACGACGATCTTATGTTCCCAATAGACCTCTGCG 2005
584 AsnThrTyrSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAl 600
2006 GGTCTCTTTGTCATCAGCTCTCTATCCAGAAATCATGTAATATAGTAG 2055
600 antiSGLyGlyAlaTyrLeuPheGlyThrTrpLysSerAlaValSerAsnL 617
2056 CCAATATCTTATGTCAGAGAGGGAGCTGCGGAGCTGGAATTTGCTAAT 2104
617 euphetTyrAlaHisAspSerSerGlyLysProIleAspAsnThrPheHis 633
```

```
2105 ..TTCCTACATAGAGATAAATTAATGAG.....CACGGCTATCGCCAT 2146
634 ArgSerLeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHI 650
2147 AGCGGTGGGTATCTGTGGGAGTTGGACCTCATCTCTTTTCTGTAGGC 2196
650 sSerPheCysLeuAlaIaGlnLeuLeuGlyLysSerSerSerP 667
2197 TACGATAAATCGGCTTTTTCGAGCTCTTCATAGAGATTAACACTAAG 2246
667 helIeThrSerThrGluThrThrSerTyrIleAlaThrVal..... 680
2247 TACATCCAAATATCATGAGACTAGTACTCAGGGGCTGTATTCTTGAG 2296
681 ..... 681
2297 GATACCTAGAGTTAGAAATGCACAGGAGTTCTATCTGATACCTCTTC 2346
681 nAlaGlnLeuAlaThrSerLeuMetLysIleSerAlaGlnAlaCysTyrA 698
2347 AGAAGCTTGCTGTACCAAGTCGTCACTATAGATATGAGTGTCTTACA 2396
698 sngLysSerIleHisGluLeuLysThrLysTyrArgSerPheSerLysGlu 714
2397 GCCATAGAAATATGATATGAAACCAAAATACAGACATAT...CCAGAA 2443
715 GlyPheGlySerThrPheHisSerValAlaValSerGlyValCysAlaSe 731
2444 GCTACGAGATCTTGGGCAATATGATGTTTGTGCTTGAGTTGAGAGCAC 2493
731 rIleProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIleP 748
2494 TACATACACTACACCTTACATCTTTTATTTGATACACTCTCGT 2543
748 heserLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSer 764
2544 TTCACAGGCTCTCAGACGACCTATGCTCACAGAAAGACTTCAAGAGACA 2593
765 SerGlyGluIleArgSerPheSerAlaSerSerPheArgAspIleSerLe 781
2594 GGAGGTGAGTTTCGCTACTTACGCGGAATCTTTCAATTTAGCAGT 2643
781 uProIleGlyIleThrPheGluLysSerGlnLysThrArgThrTyrT 798
2644 TCCATTTGGCGTGAAGTTTGAGAGATTTTCAAGCTGAAGGGAGATCT 2693
798 TyrTyrPheLeuGlyAlaTyrIleGlnAsp...LeuLysArgAspValGlu 813
2694 ATGAACCTTACCCCTTCATATGTTCTCTGATGTCATTCGCAAGATCCCAAG 2743
814 SerGlyProValLeuLeuLysAsnAlaValSerTrpAspAlaProMe 830
2744 AGC.....ACGGCAACATTGGCTAGTGAAGCTAGGTGAGACACCCACGG 2787
830 tAlaAsnLeuAspSerArgAlaTyrMetPheArgLeuThrAsnGlnArgA 847
2788 AAACAATCTCTCCAGACAGAGATTACACTGCGTTAGCAACCACTGTC 2837
847 lAlaLeuHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArg 862
2838 TCATTAATCTCTGAGTATGAGGTGTTCACTCAACGACGATTAATGATGGC 2887
863 GlyGlnSerHisSerTyrSerLeuAspLeuGlyThrTyrTyrArgPhe 878
2888 GATCTCTCTGTAATTAATTAACATCAATCTCGGGGGTAAATACCGATT 2935
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.A30847
seq_documentation_block:
ID A30847 standard: DNA; 2950 BP.
XX
AC A30847;
XX
```


1024 GATVAGTGTGACATTCCTTTTACGGGAATACAGTCACTTCTACTACTCC 1073
324 nglyLeuValArgAsnAlaIleTyLeuGluLysSerAlaIleLeuSer 341
1074 TGGGACGAAATGAACTACTATCGATTTAGAACGAGTGCMAAGATGACAG 1123
341 erLeuGluAlaArgAsnGly...AspIleLeuPhePheAspProIleVal 356
1124 CTTTCGCTTGTGCTGCTGTAGAGCACTCTACTCTATGATCCCACTACT 1173
357 GluGluSerSerSer...LysGluSerPr 365
1174 ACAGAGATCTTCACACAGATTACAGATCTCTAAAGATTAAAGACATCC 1223
365 oLeuProSerSerLeuGluAlaSerValThrsProThrProAlaThrA 382
1224 GGCAATTTCTGCACCTACAAATATACA... 1248
382 lAserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 398
1249GGGACATCATCTTTCACA 1266
399 SerGluArgLeuSerGluGluGluLysThrProAsp...AsnLeuThsE 414
1267 CGAGAAACTTATCAGACAGACAGCCCGAGATTCTTAAAAATCCTACATC 1316
414 rGlnLeuGlnGlnProIleGluLeuLysSerGlyArgLeuValLeuLysA 431
1317 GAAGTACTACAGCCTGTACTCTTTCAGAGAGTACTCTATCTTTAAAC 1366
431 sPArgAlaValLeuSerAlaProSerLeuSerGluAspProGlnAlaLeu 447
1367 ATGGAGTACCTCGACACTCAGCGATTCTCAACACGACGATCTCTCGT 1416
448 LeuIleMetGluAlaGlyThrsSerLeuLysThrSerSerAspLeuLysLe 464
1417 CTCGAAATGAGCTAGAGACTACTTAGAACCTGCTCATACTACGACCAT 1466
464 uAlaThrLeuSerIleProLeuHisLeuAspThrGluLysSerValT 481
1467 AATCAATTTGGTCATTACATCACTGCTCTATAGACCGTGCMAAGAGCA 1516
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
1517 AATATGAACCAAAAGCTAGCTCAAAAATCTGACTTTATCTGGAACCATC 1566
494 LeuSerAsnSerGlyAspGluAsnPheTyrgLysnValGluLeuLeuSe 510
1567 ACTTTATTGGACCCGACCGGACGCTTTATGAAAAATCATACTTTAAGAA 1616
510 rLysGluGlnAsnAsnIleProLeuLeuThrsLeuSer...LysGluGln 526
1617 TCCTCAGTCCAGCATCTTAGAGCTCAAAAGCTTCTGGAAGCTGACAA 1666
526 erHisLeuHisLeuProAspGlyAsnLeuSerSer.....HisPheGly 540
1667 GCACGCGAGTACCTCCAGATCTTAATAGGTGAGAAATTCATTAACGCG 1716
541 TyrgInGlyAspTrp...ThrsPheSerTrpLysAspSerAspGluGlyHis 556
1717 TATCAGGGAACCTTGGGCCCAATCTTTGCG.....GGCAGCGGGCG 1757
556 sSerLeuIleAla.....AsnTrpThrProLysAsnTyrgValProHisP 571
1758 TTCTACGAGCTCAACCTCACTGAGCTAAAGTGCCTATATCTCTATC 1807
571 rGluArgGlnSerThrLeuValAlaAsnThrLeuTyrgAsnThrTyrgSer 587
1808 CCGAGCGTATCGGCTCTTACTCCCTAATAGCTTATGAAATGACATTTATA 1857
588 AspMetGlnAlaValGlnSerMetLeuAsnThrIleAlaHisGly..... 602
1858 GATATTACCTCTCTCCATTATTTATGAGACTGCAAGCAAGCGTTGCA 1907

603 .GlyAlaTyLeuPheGlyThrTrpGlySerAlaValSerAsnLeuPheT 619
1908 GGGAAACCGTGCTTT.....TGGTGCTGCTGATTACTACTTCTTC. 1950
619 yAlaHisAspSerSerGlyLysProIleAspAsnTrpHisArgSer 635
1951CAATAGGATAGTCAAAAACACGACGCGGTTCCGCATTTGAGT 1995
636 LeuGlyTyLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
1996 GCGCGTTATGTCATAGAGGAACCTACATCTTCTGATTAAGATTCT 2045
652 ecysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerSerPheIleT 669
2046 TAGTCTGCTGATTTGTCTGAGCTCTTGGAGAAGATGAGACTACTTTGAG 2095
669 hrSerThrGluThrThrsTyrgIleAlaThrValGlnAlaGlnLeuAla 685
2096 CTAAGAAATCAAGTACAGCTTACGAGAGACCTTATTACAGACACAC 2145
686 ThrSerLeuMetLysIle..... 691
2146 GAACCTTATATCTCTCTTCCTGCAAACTACGCGCTTGTGCTGTCTTA 2195
692SerAlaGlnAlaCysTyrgAsnG 699
2196 TGTTCCTACAGAGATCTCTGTTCTTTTCAGAAACCTTGACTACACCC 2245
699 luserIleHisGluLeuLysThrLysTyrgArgSerPheSerLysGluGly 715
2246 ATACGATTAACAGATCTGAAACCAAGTATACACATATACACTTAA 2295
716 PheGlySerTrpHisSerValAlaValSerGlyGluValCysAlaSerII 732
2296 ...GGAACCTGGGGAATGATAGTTTCGCTTGAATTCGGTGAAGAGC 2342
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheS 749
2343 TCCGATTGCTTAGATGAAGTGCCTATTTGAGACGATACATGCCCTTCA 2392
749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGluGluSerSer 765
2393 TGAATTCGACTTGTCTATCCACATCAGCAAGCTTTTAAAGAACAGCGA 2442
766 GlyGluIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuPr 782
2443 ACAGAGCTCTGTGAATTTGGAAGTAGCCGCTGTGTGAATCTTGCCTTAC 2492
782 oIleGlyIleThrPheGluLysLysSerGln...LysThrArgThrTyrgT 798
2493 TATCGGATCCGATTTGATTAAGATCAGACTGCCMAAGATCMAAGCTACA 2542
798 yTyrgPheLeuGlyAlaTyrgIleGlnAspLeuLysArgAspValGluSer 814
2543 ATCTAATCTTGTGT...TATCTGTGAGTCTTGTTCGTAAGTAACCCGAC 2589
815 GlyProValIleLeuLeuLysAlaValSerTrpAspAlaProMetAl 831
2590 TGTACGACAAACACTGCGAATTAGCGGTGATTCCTGGAAGAACCTTGC 2639
831 AsnLeuAspSerArgAlaTyrgMetPheArgLeuThrAsnGlnArgAla 848
2640 GAATTTGGCAAGACAGCTTATAGTCTCTCGTCGACGAGACATTTTGTCT 2669
848 euHis...ArgLeuGlnThrLeuLeuAsnValSerCysValLeuArgGly 863
2690 TTAACCTCAAAATTTGAACCTTGTAGCCCAATTTCTTTGAAATTCGCGT 2739
864 GlnSerHisSerTyrgSerLeuAspLeuGlyThrThrTyrgArgPhe 878
2740 TCACTCCGCAATTAACATGTAGACTTAGAGACCAAAATACCAATTC 2784

[illegible]

486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerg 498
 99719 AAGTAAAAATGACCCCTTAGTGGTCCGACATTCCTTTTGGATACCAAG 99768
 498 |yAspIuAsnPherArgIuAsnValIGluLeuLeuSerLysGluGlnAsn 514
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 99769 GG.....ATGCTTATGAAATCAGCAGC...TTAGGAAAACTCAAGAC 99809
 515AsnIleProLeuLeuPheThrLeuSerLysGluInSerH 527
 99810 TTTTTCATTTGTGCACCTCTCTCTGGGTACGCAACAACATACACATGT 99859
 527 sLeuHIsLeuProAspGlyAsnLeuSerSerHisPheGlyTyrGlnGlyA 544
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 99860 TCCAGGACCTTCCACAGTACGAAACCTCCTACGACACTATGGGTATCAAGCTA 99909
 544 sPrlPhePheSerTrpLysAsp.....SerAspGluGlnHis 556
 99910 CTGGGGGATGACTTGGTTGATGATACCCGACGACCTCCAAAGACTAG 99959
 557 SerLeuIleAlaAsnTrpPheProLysAsnTyrValProHisProGluAr 573
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 99960 ACAGGCACATTAAGCTTTGGACCAATACAGGCTACCTTTCCGATCTTACGC 100009
 573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
 100010 TCCAGGACCTTTAGTCCCAATACCTTTGGGATCTTTTTCAGACATCC 100059
 590 InAlaValGlnSerMetIleAsnThrIleAlaHisGlyGlyAlaTyrLeu 606
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100060 AAGCATTCATACGTCATAGAGAGAGAGCTTGGACTCTTGGTTCAGAT 100109
 607 PheGlyThrTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspS 623
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100110 CCAGGCTTTCGGCTCGCGGAGTCCGCAAT...TTCTTGATTAAGATTA 100156
 623 rSergLysProIleAspAsnTrpHisIleArgSerLeuGlyTyrLeuP 640
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100157 GAAGGCGGCAAAA...CGCAATATACCGCTATTAATCTGGHGATAGTCA 100203
 640 hGlyLysSerThrHisSerLeuAspHisSerPheCysLeuAlaIle 656
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100204 TCGGAGGTGCACGCAACCTTGTTCGAAAACTAATTATGCTTGGCTTT 100253
 657 GlyGlnIleLeuGlyLysSerSerAspSerPheIleThrSerThrGln 673
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100254 TCCCAACTTTCGTGATGCATTAACGATTTCTTAGTGGCTAAAAATCATAC 100303
 673 rThrSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100304 TGCATACATATGCAGAGAGCCTTATATCCAACACATTAACGATGTAGTG 100353
 688 688
 100354 GGTTCATAGTGTCTCTTAGATAAACTCCGCGCTTGGAGTCAATAA 100403
 689 ...MetLysIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGlu 704
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100404 CCCCTCGTTTACAAAGGAGCGCTCGCTTATATGGCACGCTAGTATATATCT 100453
 704 uLysThrLysTyrArgSerPheSerLysGluGlyPheGlySerTrpHis 721
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100454 GAAGCAAAAGCTACTGCCTAT...CTGATGGAAAGCTTCTTGGGGCA 100500
 721 eValAlaValSergLysGluValCysAlaSerIleProIleValSerAsn 737
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100501 ATAAAGCTTTTACATGATGTTGGGAGCTTCTTCATCTTATCTCTGAA 100550
 738 GlySerGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyH 754
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 100551 TACCGCATGTGTTTGATACCTATGCTCATATCAATCAACTGATCTGAC 100600
 754 eSerGlyHrgIuAsnArgLysPheGluGluSerSerGlyGluIleArgSer 771

223 ThrGlyAsnSerAlaThr.....As 229
|||:|||||
833 ATTGACATAACCTCAGACGACGACCTCCACACACCGGATATGTC 882
229 nGlyAlaIleCysCysIleSerAspLeuAsnThrSerGlyGlyS 246
:|||||:|||||
883 AGGAGTGTCTCTGCT.....TATAACTAGTACGATCTA 923
246 erLeuSerLeuAlaCysAsnGlnIleThrLeuPheAlaSerAsnSerAla 262
:|||||:|||||
924 AGGTACCCCTCAGTCAATCAGATGTTACTCTTACGACAAATACATCG 973
263 LysGlyLysGlyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAs 279
|||||:|||||:|||||
974 ACAACAGCGGAGGAGCATCTATCTATGAAAAAGCTCGAAGCTGGCTCCG 1023
279 nGlyProValSerPheIleAsnAsnSer.....AlaLysI 291
|||:|||||
1024 AGGACTTACCCCTATTCAGTAAATAGTGTCAATGAGGATACAGCTCTCA 1073
291 IeGlyAlaIleAlaIleAlaIleGlnSerGlySerLeuSerIleLeuAla 307
|||||:|||||:|||||
1074 AAGGTGAGCCATGACTATCGAAGATAGTGGGAATGAGTTATTCGCC 1123
308 GlyGlyLysSerValLeuPheGlnAsnAsnSerGlnArgThrSerAspG 324
:|||||:|||||
1124 CATACTGGTGCATCTCTCTTTTACGCAATACAGTCACCTTCTACTAC 1173
324 nGlyLeuValArgAsnAlaIleTyrLeuGlyLysAspAlaIleLeuSerS 341
|||:|||||:|||||
1174 TGGGACGATGAGATAGTATGATGAGAACGAGTCAAGATGACAG 1223
341 erLeuGlnAlaArgAsnGly...AspIleLeuPhePheAspProIleVal 356
:|||||:|||||
1224 CTTCGTTCTGCTGCTGTGAGAGCATCTCTCTATGATCCATTA 1273
357 GlnGlySerSerSer.....LysGlySerPr 365
|||:|||||
1274 ACAGATCATCTCCACACAGTACAGATGCTCTTAAAGTTAAACAGAC 1323
365 OLeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrA 382
-|||:|||||
1324 GGCAGATTCTGCACACTACAAATATACA..... 1348
382 IAserProLeuValIleGlnIleThrSerAlaAsnArgSerValIlePheSer 398
1349GGGACATCATCTTCCACA 1366
399 SerGluArgLeuSerGlyGlyLysThrProAsp...AsnLeuThrSe 414
:|||||:|||||
1367 GGAGAAAGTTATCTAGACACAGACGCCGCCAGATTCTAAAGTCTACTTC 1416
414 rGlnLeuGlnIleProIleGlyLeuLysSerGlyArgLeuValLeuLysA 431
:|||||:|||||
1417 GAAGCTACTACAGCCTGTAACCTTTCAGAGAGTACTATCTTAAAC 1466
431 spArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 447
1467 ATGGAGTACTCTGCAGACTCAGGACTTCACTCAACAGCAGATCTCTG 1516
448 LeuIleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLe 464
|||:|||||:|||||
1517 CTCGAAATGACAGTACAGCACTACTACACCTCTCATACAGACCAT 1566
464 uAlaThrLeuSerIleProLeuHisSerLeuAspThrGlyLysSerValT 481
:|||||:|||||
1567 AAACAATTGTGTCATTACATCAGTCTATAGACGCTCAAAAGAGCGCA 1616
481 hrIleHis.....AlaProAsnLeuSerIleGlnLysIlePhe 493
|||:|||||
1617 AATATGAACCAAGCTACGTCAAAAAATCTGACTTTATCTGGAACCATC 1666

494 LeuSerAsnSerGlyAspGlnAsnPheTyrGlnAsnValGlyLeuLeuSe 510
:|||||:|||||
1667 ACTTATTCGACCCGACCGGACGCTTTTATGAAATCATAGTTTACGAA 1716
510 rLysGlnGlnAsnAsnIleProLeuLeuThrLeuSer...LysGlnGlnS 526
:|||||:|||||
1717 TCCTCAGTCTCAGCATCTTAGAGCTCAAGCTTCTGGAAGTGAACA 1766
526 erHisLeuHisLeuProAspGlyAsnLeuSerSer.....HisPheGly 540
|||:|||||
1767 GCACCCGAGTCACTCCAGATCCCTAATGCGTGAAGAAATTCATACGCG 1816
541 TyrGlnGlyAspTrp...ThrPheSerTrpLysAspSerAspGlyHis 556
|||||:|||||
1817 TATCAGGCACTTGGGCCCAATGTTG.....GGACAGGGGC 1857
556 sSerLeuIleAla.....AsnTrpThrProLysAsnTyrValProHisP 571
|||:|||||
1858 TTCTACAGACTGCACACTCACTCACTGACTAAAGCTGATATTCCTATC 1907
571 rGluArgGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSer 587
|||||:|||||
1908 CCGAGCGTATCGCGCTTACTCCCTAATAGCTATGGAATGCAATTTATA 1957
588 AspMetGlnAlaValGlnSerMetIleAsnThrIleAlaHisGly..... 602
:|||||:|||||
1958 CATATTACTCTCTCATTAATCTTATGAGACTGCAAGAGGCTTGCA 2007
603 GlyAlaTyrLeuPheGlyTyrThrGlySerAlaValSerAsnLeuPheT 619
|||:|||||
2008 GGAAGACCGTGCTTT.....TGGTGTGCGATATCTACTTCTCTC 2050
619 yTrAlaHisAspSerSerGlyLysProIleAspAsnTrpHisArgSer 635
|||:|||||
2051CATAGAGTATGACAAAAACACGACGCGGCTTGCCCATTTGACT 2095
636 LeuGlyTyrLeuPheGlyIleSerThrHisSerLeuAspAspHisSerPh 652
|||||:|||||
2096 GCGGTTATGTCATAGAGGAAGAACTACATACTTCTAGATTAAGATTCT 2145
652 eCysLeuAlaIleGlyGlnLeuLeuGlyLysSerSerAspSerPheIleT 669
|||:|||||
2146 TAGTCTCATTTTGTCTGAGCTCTTGAAGAGATGAGACACTTGTAG 2195
669 hrSerThrGluThrThrSerTyrIleAlaThrValGlnAlaGlnLeuAla 685
:|||||:|||||
2196 CTAAAGATCAAGGTACAGTACGAGGAGACTCTATTAACACGACAC 2245
686 ThrSerLeuMetLysIle.....SerAlaGlnAlaCysTyrAsnG 691
:|||||:|||||
2246 GAACCTTATCTCTCTCTCTTCCAAACTACGCGCTTGTCTGTCTTA 2295
692SerAlaGlnAlaCysTyrAsnG 699
2296 TGTTCCTACAGAGATTCCTGTTCTCTTTACGAAACCTTAGCTACACC 2345
699 IuSerIleHisGlyLeuLysThrLysTyrArgSerPheSerLysGlyGly 715
:|||||:|||||
2346 ATACGATTAACGATCTCAAAACCAAGTATACACATATCACTACTGTA 2395
716 PheGlySerTrpHisSerValAlaValSerGlyGluValLysAlaSerI 732
|||||:|||||
2396 ...GGAACCTGGGGAATGATAGTTTCCCTTACAAATTCGTCGAAGAC 2442
732 eProIleValSerAsnGlySerGlyLeuPheSerSerPheSerIlePheS 749
|||||:|||||
2443 TCCGATTGTCTAGATGAAGTCTCTATTGAGCAGTATGCCCTTCA 2492
749 erLysLeuGlnGlyPheSerGlyThrGlnAspGlyPheGlnGlyLysSer 765
|||||:|||||
2493 TGAATTTCAGTGTGTCTATGACATCAGGAAGTTTAAAGACAGGGA 2542
766 GlyLysIleArgSerPheSerAlaSerSerPheArgAsnIleSerLeuP 782

189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAl 205
189 ValSerProIleSerPheAlaArgAsnArgAlaAspLeuAsnGlyGlyAl 205
643 ACCGGCAGGAGTACTTATGATATATTAAGAGAACTGGGGTGGGCC 692
205 aIleCysCysSerAsnLeuIleCysSerGlyAsnVal.....Asp 219
693 TCTC.....GGCTTGAACCCACCTCTCCATTAACCAAAATACCT 733
219 rolePhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
734 CCTTTTCTCTCTGGAAACACTGCACACAGATGCTGCAGGCAAGGGCGG 783
232 AlAlIleCysCysIleSerAspLeuAsnThrSerGlyGlySerLeuSe 248
784 GCCATTATATGT.....GAAAAACAGAGAGAGACTCTACTACTTAC 824
248 rLeuAlaCysAsnGlnGluThrLeuPheAlaSerAsnSerAlaGlyGlu 265
825 TATCTCTGGAAATAAACTCTGACCTCCCGAAGACCTTCACTAATC 874
265 ySGlyGlyAlaIleTyrAlaLysIleMetValLeuArgTyrAsnGlyPro 281
875 AAGGGGAGACATCTGTGCCATGTCTAGATCTTCCGCTGTGGCCCT 924
282 ValSerPheIleAsnAsn.....SerAlaLysIleGlyAl 293
925 ACCCATTTTCAATTAATAGATGGCGAGACACACTGCAAGCAAGGGCGG 974
293 yAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGluG 310
975 CCTATTGCAATTCGCCACTCTGCATCTTAACTCTCTCAAAATCAAG 1024
310 ySerValLeuPheGlnAsnAsnSer...GlnArgThrSerAspGlnGly 325
1025 GAGCATATACGTTCTTGGCAACACTTAACCTCAACCTCCGCGCAACA 1074
326 LeuValArgAsnAlaIleTyrLeuGlnLysAspAlaIleLeuSerSerI 342
1075 TCGACACGAGATGCTATCTACTGCGATCGTCACCAAAATTAACAACCT 1124
342 uGluAlaArgAsnGly...AspIleLeuPhePheAspProIleValGlnG 358
1125 AAGGGCACCACCAAGCCCAATCTATCTTATGATCCGATTCATCTCA 1174
358 luserSerSerGlySerProLeuProSerSerLeuGlnAlaSerVal 374
1175 ACACCAAGAGACTTCAGACGTTCTG..... 1200
375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
1201 ACCATCAACCAACCGATAGCAACTCGCCTTA.....GATTA 1238
391 AsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGluL 407
1239 TTCAGCAGACGATTTGATTTTGGGGAAAGCTCTCTGCAGATVCAAGCA 1288
407 ySThrProAspAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
1289 AAGCTGCTGATACTTACATCTATATTAAGCAACCATTTGGCTCTAGCC 1338
424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerI 440
1339 TCTGGACCTTAGCACTCAAGAAATGCGATAGATGTCATAGCTT 1388
440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeuL 457
1389 CACAGACAGCTGAAGCTCTACACTCTCTCAACACGAGCAACAAGCTCA 1438
457 ySThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
1439 AAGCAATATCTAAGCTATCACTTACCAAACTGTCTGTTGATCTTCT 1488
473 SerLeuAspThrCluLysSerValThrIleHisAlaPro..... 485

1489 GCCCTTAAGAGAAATAAGACTGTGTCATTAAGAAACACAGGACCAACA 1538
486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 500
1539 AACTATATCTTAACTCTCTCTCTTTTTCAGAGATAGTAGGGC.... 1584
500 LysAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsnAsnI 516
1585 ..AATTTTATGAAGC...CATACGATAAACCAAGCCTTCAGCAGACCT 1629
516 e.ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAsp 532
1630 TTGCTGTATTAATACGCTGCTACTGCTCTGACGATATATTAATC...GAT 1676
533 GlyAsnLeuSerSer.....HisPheGlyTyrG 542
1677 GCGCTTCTCACTCTCCAGTACAACTCCAGACCTCATTAAGGATATCA 1726
542 nGlyAspTrpThrPheSerTrpLysAspSerAspGluLysIleSerLeuI 559
1727 GGCACATTTGGAAACCACTTGGCCAGACATCACTCAAAATCAAGAA 1776
559 lAlaAsnTrpThrProLysAsnTyrAlaProHisProGluArgGlnSer 575
1777 CTATCACTTGGCTAATCACTGAGCAACCTTAATCTGACCGTAGAGCT 1826
576 ThrLeuValAlaAsnThrLeuThrAsnThrTyrSerAspMetGlnAlaVal 592
1827 TCCGTACTCTCCGATTCATTAATGGGCATCTTATGATACATTCACCTCT 1876
592 lGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGlyT 609
1877 ACACAGAGATCATGACATCTCAAGCAATAGATATACAGCAAGAGAGAC 1926
609 hTrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGly 625
1927 TCTGGCATCAAGAACTCGCAATTTCTTCAT...AAGCAATAATTCAGGA 1973
626 LysProIleAspAsnTrpHisIleArgSerLeuGlyTyrLeuPheGlyI 642
1974 ACTAAC...CAAGCATTCGCAATATAAAGCTACGCTATATTTGAGG 2020
642 eSerThrHisSerLeuAspAspHisSerPheCysLeuAlaIleGlyGln 659
2021 AAGTCTGATGATTTTCTGAAATATCTTCAGTACCTTTCTGCGAGC 2070
659 eLysGlyLysSerSerAspSerPheIleThrSerThrGluThrThrSer 675
2071 TCTTCGGAAGATTAAGACCTGTATTATGTAAGAAATACCTTCATAG 2120
676 TyrIleAlaThrValGlnAlaGln..... 683
2121 TATTACCGCTGCTATTAACCTGCACATCGACATTCCTAGAGAGACTTCC 2170
684LeuAlaThrSerLeuMetLys..... 690
2171 CATGCCCTCATTTGGAACTATCACAGACATGCTGAAAGATATTCCTGCA 2220
691 ..IleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeuLysThr 706
2221 TTTCGAAAGCCAGCTGAAGCTACACTCACTAAAGAAATAGATGATCT 2270
707 LysTyrArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAl 723
2271 CGCTTACTTCTTAT...CTGAAAGCTCAAGCCTTTTGACCAATTAATC 2317
723 ValSerGlyGluValCysAlaSerIleProIle...ValSerAsnGlyS 739
2318 TGGGCTCTAGAGCTCGAGAGATCTCTGCTATATCTCCCTAAAGAG 2367
739 eGlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSer 755

471 LeuH1SerLeuAspThrGluYSerValThrIleHisAlaPro.... 485
 1477 GTAGACTCTTTAAGCGGAGGATGAAGAAATTTGATTTGCTCTTCGACG 1526
 486AsnLeuSerIleGluLysIlePheLeuSerAsnSerG 498
 1527 AAGTAAATATGAGCCCTTATGTCGCGCATCTCTTTGGATATACCAAG 1576
 498 LysAspGluAsnPheTyrGluAsnValGluLeuLeuSerLysGluLysn 514
 1577 GG.....AATGCTTATGAATAATCCAGC...TAGAGAAATATCAAGAC 1617
 515AsnIleProLeuLeuThrLeuSerLysGluGlnSerH 527
 1618 TTTTCATTTGTCCACCTCTGCTGCTGGGTACTGCAACACATACAGATGT 1667
 527 sLeuH1sLeuProAspGluAsnLeuSerSerHisPheGlyTyrGlnGly 544
 1668 TCCACGCGGTCTACAGTAGACACACTCCATACCCACTATGGATACAGTA 1717
 544 sPTrPTrPheSerTrpLysAsp.....SerAspGluLysHis 556
 1718 CTGGGGGAAATGACTGGGGTGTGATGATACCGACAGACTCCAAAGATAAG 1767
 557 sSerLeuIleAlaAsnTrpThrProLysAsnTyrAlaProHisProGluAr 573
 1768 AAGGCGACATTAAGCTTTGAGCAATACAGCAACCTCCGAAATCTTGAGCG 1817
 573 gGlnSerThrLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetG 590
 1818 TCAGAGACTTATAGTCTCTAATAGCTTTGGGGATCTTTTCAGACATCC 1867
 590 IAlaValGlnSerMetIleAsnThrIleAlaHisGlyAlaTyrLeu 606
 1868 AAGCATTCACAGGTGTCTACAGAGAGAGTGGCTTGACTCTGTTCAGAT 1917
 607 PheGlyThrTrpGlySerAlaValSerLysnLeuPheTyrAlaHisAspSe 623
 1918 CGAGGCTTCTGGGCTGCGGGATCCCAAT...TTCTAGATAAAGATTA 1964
 623 rSerGlyLysProIleAspAsnTrpHisIleArgSerLeuLysTyrLeuP 640
 1965 GAAGAGGAGAAA...CCCAATATCCGTCATTAATCTGTGGATATGCTA 2011
 640 hGcIlyIleSerThrHisSerLeuAspAspHisSerPheCysLeuAla 656
 2012 TCGGAGGTGCACGCAAACTGTCTGCAAACTTAATAGCTTGGCTTT 2061
 657 GlnGlnLeuLeuGlySerSerSerSerPheIleThrIleSerTrpGluTh 673
 2062 TCCCAACTCTTTGGTAGCGATTAACAATTTCTTAGTCGGTAAATAATCAAT 2111
 673 rHisSerTyrIleAlaThrValGlnAlaGlnLeuAlaThrSerLeu... 688
 2112 TCAATACCTATGAGAGAGCTTGTATATCCACACATTTACAAATATGATG 2161
 688 688
 2162 GGTTCATAAGTGTCTTAGATAAATCTCTGAGCTTGGAGATCAATAA 2211
 689 ...MetLysIleSerAlaGlnIleCysTyrAsnGlnSerIleHisGlyLe 704
 2212 CCCCTCGTTTGAAGAGGACAGCTCCCTTATATACCAACGTCAGTATGATCT 2261
 704 cLysThrLysTyrArgSerPheSerLysGluGlyPheGlySerTrpHis 721
 2262 GAAGCAAAAGTATTACTGCGTAT...CTGAGGTGAAGAGTCTTGGGGGA 2308
 721 eValAlaValSerLysGluValCysAlaSerIleProIleValSerAsn 737
 2309 ATAATGCTTTTAACATGATGTTGGAGACTTCTTCATCTTATCTGA 2358

[illegible]

189 ValSerProIleSerPheAlaIArgAsnArgAlaAspLeuAsnGlyVal 205
 743 ACCGGCCAGTAGTCTTATGATATGAAGGAAAGTGGGGGTGGGC 792
 205 aIleCySerAsnLeuIleCySerGlyAsnVal.....Asn 219
 793 WCTG.....GGCTTGAAGCCACCTCCCTCATTTACTCAAAATACCT 833
 219 rOleuPhePheThrGlyAsnSerAlaThrAsn.....GlyGly 231
 834 CCTTTTCTCTCTGAAACATCGACAGATGTGACAGGCAAGGGCGG 883
 232 AlaIleCySerGlyIleSerAspLeuAsnThrSerGlyGlySerLeuSe 248
 884 GCCATTATTGT.....GAAAAACAGAGAGACTCTACTACTTAC 924
 248 rLeuAlaCyAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGlu 265
 925 TATCTCTGGAATAAAGCTACCTTCGCCAGACACTTCACGTAACTC 974
 265 ysglyGlyAlaIleTyrAlaLysHisMetValLeuArgTyrAsnGlyPro 281
 975 AAGCGGAGCATCTGTGCCATGTGTATGATCTTCCGCTCGCCCT 1024
 282 ValSerPheIleAsnAsn.....SerAlaLysIleGly 293
 1025 ACCCTATTTCATAAATAGATGCGGGAACACAGCTCGACGCAAGGCGG 1074
 293 yAlaIleAlaIleGlnSerGlyGlySerLeuSerIleLeuAlaGlyGln 310
 1075 CCGATTTCATTCGCCGACCTCGACCTGCACTTTAAGTCCTCTGCAAAATCAAG 1124
 310 lYserValLeuPheGlnAsnAsnSer...GlnArgThrSerAspGlnGly 325
 1125 GAGCATCATCTCTCTGCAACACTCAACCTCAACCTCCGCGCCACA 1174
 326 LeuValArgAsnAlaIleTyrLeuGluLysAspAlaIleLeuSerSerLe 342
 1175 TCGACCGGAATGCTATCTACTCTGGATGTCGACAAATTTACCACTT 1224
 342 uGluAlaIArgAsnGly...AspIleLeuPhePheAspProIleValGln 358
 1225 AAGCGACCCCAAGGCCAATCATCTATTCTCATATCCGATTCGATCTA 1274
 358 lUserSerSerLysGluSerProLeuProSerSerLeuGlnAlaSerVal 374
 1275 ACACCAAGAGGAGCTTCAGACGCTCTG..... 1300
 375 ThrSerProThrProAlaThrAlaSerProLeuValIleGlnThrSerAl 391
 1301 ACCATCAACCAACCGATAGCAACTCGCCTTA.....GATTA 1338
 391 aAsnArgSerValIlePheSerSerGluArgLeuSer...GluGluGlu 407
 1339 TTCACGACGATTTCTATTCTCGGGAAGACTCTCTGCACATGCAAGCA 1388
 407 ySThrProAspAsnLeuThrSerGlnLeuGlnProIleGluLeuLys 423
 1389 AAGTGTGATAAATCTCATATTAATAAGAACCACTTGCTGCTACCC 1438
 424 SerGlyArgLeuValLeuLysAspArgAlaValLeuSerAlaProSerLe 440
 1439 TCTGGACCTTAGCACTCAAGGAATGTGAGTTGATGTCAATGATT 1488
 440 uSerGlnAspProGlnAlaLeuLeuIleMetGluAlaGlyThrSerLeu 457
 1489 CACACAGACCTAAGGCTCTACACCTCATGCAACCAAGCAAGCAAGCTCA 1538
 457 ySThrSerSerAsp...LeuLysLeuAlaThrLeuSerIleProLeuHis 472
 1539 AAGCAGATACGGAAGCTATCAGTCTTACCAAACTGTGTGATCTTCT 1588
 473 SerLeuAspThrGluLysSerValThrIleHisAlaPro..... 485

1589 GCCTTAGAGGAAATAGAGTGTGTCTCATGAACAGCAGGACCAACAA 1638
 486AsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAsp 500
 1639 AACTATACCTCAACCTCTCTTGTGTTTCCAAATAGTAGCGGC... 1684
 500 lUAsnPheTyrGluAsnValGluLeuLeuSerLysGluGlnAsnAla 516
 1685 ..AATTTTATGAAGAACCATAGATAAACCAAGCTTACCCAGCTTGT 1732
 517 ProLeuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspG 533
 1733 GTCGATTTCACCTGCTGCTACGCTCTAGCCATATTATATC...GATCC 1779
 533 yAsnLeuSerSer.....HisPheGlyTyrGln 543
 1780 CTTCTCACTTCTCCAGTACAACTCCAGAACCTCATTAACGGATATCAG 1829
 543 lYAspTrpThrPheSerTrpLysAspSerAspGluGlyHisSerLeu 559
 1830 GACATTGGGAAGCCACTGGGAGACACATCACTGCAAAATCAAGAACT 1879
 560 AlaAsnTrpThrProLysAsnTyrValProHisProGluArgGlnSerTh 576
 1880 ATGACTTCGCTACTACGCGCTACACCTTAATCTGAGCTAGAGCTTC 1929
 576 rLeuValAlaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValG 593
 1930 CGTACTTCGCTATCATATATGCGATCTTATGACATTCGACACTCTAC 1979
 593 lAsnMetIleAsnThrIleAlaHisGlyAlaTyrLeuPheGlyThr 609
 1980 AGCAGATCATGACATCTCAACGGAATAGTATCTACGACCAAGAGACTC 2029
 610 TrpGlySerAlaValSerAsnLeuPheTyrAlaHisAspSerSerGly 626
 2030 TGGCGATCAGAACTCGCAATTTCTTCAT...AAGCATTAATCAGGAA 2076
 626 sProIleAspAsnTrpHisArgSerLeuGlyTyrLeuPheGlyIle 643
 2077 TAAC...CAACATTCGCACATTAAGCTACGCTATATTCTTGGAGCA 2123
 643 erThrHisSerLeuAspAspHisSerPheCysLeuAlaIleGluLeu 659
 2124 GTGCTGAAGATTTTCTGAATAATCTCACTGATGACTTCTGCGAGCTC 2173
 660 LeuGlyLysSerSerAspSerPheIleThrSerThrGluThrTrpSerTy 676
 2174 TTCGCTAAAGATAAAGACCTGTTATAGTGAATAACCTCTCATACTA 2223
 676 rIleAlaThrValGlnAlaGln..... 683
 2224 TTTAGCTTCCTATACCTGCAACATCGAGACTTCTTAGAGAGACTTCCA 2273
 684LeuAlaThrSerLeuMetLys..... 690
 2274 TGCCCTCATTTGGAAGTATCACGCAATGCTGAAGAATATCTCCATCT 2323
 691 lIleSerAlaGlnAlaCysTyrAsnGluSerIleHisGluLeuLysThrLy 707
 2324 TTTGATGCCAGCTAAGCTACGCTACACTAAATATGATAGTACTGCG 2373
 707 sTyrArgSerPheSerLysGluGlyPheGlySerTrpHisSerValAlaVal 724
 2374 CTATACCTTCCAT...CTGAAAGCTCAAGCTCTTGGACCAATACCTGC 2420
 724 aISerGlyLysValCysAlaSerIleProIle...ValSerAsnGlySer 739
 2421 GGGCTTAGACCTCGAGAGATCTGCTGCTATATCTCCCAAGGAAGCA 2470
 740 GlyLeuPheSerSerPheSerIlePheSerLysLeuGlnGlyPheSerCl 756
 766

```
192 leSerPheAlaIrgAsnArGAlaAspleuAnGlyAlaIleCysCys 208
      ::::::::::::::::::::
511207 TAACCTTGGCAAAACAAACACGCAAAAGGCGCGTCCCTCATTTCC 511256
      ::::::::::::::::::::
209 SerAnleuIleCysSerGlyAsn...ValAsnProLeuPhePheThrCl 224
      ::::::::::::::::::::
511257 ACGGGAGGGATTACAAATTACATACGTTAACTACGATCATTTTTCGA 511306
      ::::::::::::::::::::
224 YAsnSerAlaThrAsn..... 229
      ::::::::::::::::::::
511307 AAATACCGCGGCGAAACAATGGCGAGCCATTACACGCAAGCTAGCACTT 511356
      ::::::::::::::::::::
229 ..... 229
      ::::::::::::::::::::
511357 TTATTACGACGACAACCAATTAACCTTATTAACAATAGTTGACCCCA 511406
      ::::::::::::::::::::
230 .....GlyGlyAlaIleCysCysIleSerAspleuAsnThrSe 242
      ::::::::::::::::::::
511407 ACCTGACGTACAGGGGAGGCGATTACTAGTAGT.....ACATGAGC 511450
      ::::::::::::::::::::
242 rGluIysGlySerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAlas 259
      ::::::::::::::::::::
511451 CCCCAGACAGCTTAACTACTATCAGACACGCGCACTCACTTTATAG 511500
      ::::::::::::::::::::
259 eAsnSerAlaIysGluIysGlyAlaIleTyAlaIasHisNeVal 275
      ::::::::::::::::::::
511501 CAAATACACCAATTAATCTACTGCTGGCGCATTTATACATCATCTACT 511550
      ::::::::::::::::::::
276 leuATyTyraSncIyProValSerPheIleAsnAsnSer..... 288
      ::::::::::::::::::::
511551 CTCTCTCTGGAGACTAGCTCTTTTAAACAACCTGCTGCTATAGATAC 511600
      ::::::::::::::::::::
289 .....AlaIysIleGlyAlaIleAlaIleCysIleSerGlySerLeuS 304
      ::::::::::::::::::::
511601 TCGAGCTCCCTTAGACAGACCAATTCGATTCGACATCTGATCTTTCA 511650
      ::::::::::::::::::::
304 erIleLeuAlaGlyGluIysSerValLeuPheGlnAsnAsnSer..... 318
      ::::::::::::::::::::
511651 GTCCTTCGCGCTCTGCTGACACATCACTTGGAGCAACACATAGTAC 511700
      ::::::::::::::::::::
319 ...GlnATyThrSerAspGlnIyLeuValArGAsnAlaIleTyLeu... 333
      ::::::::::::::::::::
511701 AAGGAGCTTCTTGGAGTCAACCACTACAGAAATTTCTATTACATCGG 511750
      ::::::::::::::::::::
334 GluIysAspAlaIleLeuSerSerLeuGluAlaArGAsnIyAsp... 349
      ::::::::::::::::::::
511751 AAACACCAATCTAAGATTGTACAGCTCCAGCCCTCAACCCATTACTA 511800
      ::::::::::::::::::::
349 leLeuPheAspProIleValGlnIysSerSerLeysGluSerPro 365
      ::::::::::::::::::::
511801 TCTACTTCTAGATCCATTAACA..... 511823
      ::::::::::::::::::::
366 leuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
      ::::::::::::::::::::
511824 .....ACTACATCATCTGACAGCTCTCAAGTGTCTTAACCTTAAGG 511867
      ::::::::::::::::::::
382 aSerProLeuValIleGlnThrSerAlaAsnArGSerValIlePheSers 399
      ::::::::::::::::::::
511868 TCCGTACCTTCAGAGGAAATCCCATATCAAGAACCAATCCATTTTCG 511917
      ::::::::::::::::::::
399 erGluArGLeuSerGluGluIySthr...ProAspAsnLeuThrSer 414
      ::::::::::::::::::::
511918 GAGAGAACCTCTCGAGACCAACCTCGAGAAAGCATATATCTCAAACT 511967
      ::::::::::::::::::::
415 GlnLeuGlnGlnProIleGlnLeuIysSerGlyArGLeuValLeuIySAs 431
      ::::::::::::::::::::
511968 ACAATTCAGCAACCTCTACTCTTGGGAGGAGCACTCTCTTTAAATC 512017
      ::::::::::::::::::::
431 pArGAlaValleuSerAlaProSerLeuSerGlnAspProGlnAlaLeuL 448
      ::::::::::::::::::::
512018 AGGAGTCACTCACTAGTTCGCAAGTCTTTTCGAAATCCCGGCGCTACCC 512067
      ::::::::::::::::::::
448 euIleMeGluAlaGlyThrSerLeuIySthrSerSerAspleuIySleu 464
      ::::::::::::::::::::

      ::::::::::::::::::::
512068 TCCCTATGATGACGAGGACCACTTAGAAAC...GGTGAATGATCATTA 512114
      ::::::::::::::::::::
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluIysSerValTh 481
      ::::::::::::::::::::
512115 TCAATTAATCTTCTCAATGTATGATTCCTTAAAGAGCCAGAAAGTATC 512164
      ::::::::::::::::::::
481 rIleHisAlaProAsnLeuSerIleGlnIySlePheLeuSerAsnSer 497
      ::::::::::::::::::::
512165 CCTAAACCAACACACAGCAACT...CAGACAGTCACTTATCTGATGCC 512211
      ::::::::::::::::::::
498 .....GlyAspGluAsnPhetyrGluAsnValGluLeuLeu 509
      ::::::::::::::::::::
512212 TCTCTCTGTAGATCCTTGTGAATGTCTAGAGATGC..... 512252
      ::::::::::::::::::::
510 SerIySGluGlnAsnAsn.....IleProLeuLeuThrLeuSerly 523
      ::::::::::::::::::::
512253 .....TCTTGAATACCTCAAGCTCTTTCTGTCTCACTCTTACTGC 512296
      ::::::::::::::::::::
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
      ::::::::::::::::::::
512297 TCACGACCCCGCGAATATTCACATCAGAC.....TTAGCTCTGATC 512340
      ::::::::::::::::::::
538 .....HisPheGlyTyrgInIyAspTrpThrPhe 547
      ::::::::::::::::::::
512341 CCCTAGAAAAAATCTATCATCTGAGGATACCAAGGAATTTGGGCACTTA 512390
      ::::::::::::::::::::
548 SerTrpIyAspSerAspGluGlnHisSerLeuIleAla.....AsnTr 562
      ::::::::::::::::::::
512391 TCTTGGCAAGGATCTCGACTAAATCCAAGCAGGACTCTTACTCG 512440
      ::::::::::::::::::::
562 TrpThrProIyAsnTyraValProHisProGluArGlnSerThrLeuVal 579
      ::::::::::::::::::::
512441 CACAAAAACAGATATACATCCGAAATCCGACGTCGTGACACCTTAGTTC 512490
      ::::::::::::::::::::
579 laAsnThrLeuTrpAsnThrTyTrSerAspMetGlnAlaValGlnSerMet 595
      ::::::::::::::::::::
512491 CTACACACTATGGGATCTTGTGTGATGTGGCTCATCAACAGCTT 512540
      ::::::::::::::::::::
596 IleAsnThrIleAlaHisGlyAlaValArTyLeuPheGlyThrTrpGlySe 612
      ::::::::::::::::::::
512541 GTAGCCACTAAGTACGCCAATCTCAAGAACTCGCGGACTGTGTGTA 512590
      ::::::::::::::::::::
612 rAlaValSerAsnLeuPheTyraIleHisAspSerSerGlySproIleA 629
      ::::::::::::::::::::
512591 AGGATCTCGACACTTCTC...CATAAAGATACGACGAAGTAATA 512634
      ::::::::::::::::::::
629 sPAsnTrpHisHisArgSerLeuGlyTyTrLeuPheGlyIleSerThrHis 645
      ::::::::::::::::::::
512635 AAGGTTTCGCCACATAGTGCAGGTTATGTGTAGAGGCACTACACA 512684
      ::::::::::::::::::::
646 SerLeuAspAspHisSerPheCysLeuAlaIaGlyGlnLeuLeuGlyLy 662
      ::::::::::::::::::::
512685 TTAGCTTGTGAATATTATATCATCTGCAGCTTGTGCCAATTATTCGGGA 512734
      ::::::::::::::::::::
662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyIleAlaI 679
      ::::::::::::::::::::
512735 ACATAGACATCATCTTAATAATAATAATAGACCTTCCTCATATCAGCTT 512784
      ::::::::::::::::::::
679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetIyS 690
      ::::::::::::::::::::
512785 CTCTCATCTCCACATCTACGACCTTGTCTCTCCAAAGCTTGTACGC 512834
      ::::::::::::::::::::
691 .....IleSerAlaGlnAl 695
      ::::::::::::::::::::
512835 TACCTTCTGATCTGAAGTACAGGCTTCTCTTTGATGCTCAAT 512884
      ::::::::::::::::::::
695 acTyTyraSngIySerIleHisGluLeuIySthrLysTyraSrgSerPheS 712
      ::::::::::::::::::::
512885 CAGCATATCTATAGTAATAATATAGATGAACCATATACACCAAGAC 512934
      ::::::::::::::::::::
712 erLySGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
      ::::::::::::::::::::
```


159 HisAsnTyrLeuPheThrAlaPheGluGluAsnSerSerLysGlyAsnGlu 175
||||| ||| ::||| ||||| ::|||
1409 TCGAACTAT...AGTTCCTACTTGGCCAAACACTTTCTTAATGACAAATCG 600
175 YGlyAlaIleGlnAlaGlnThrPheSerLeuSerArgAsnValSerPro 192
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
601 AGCCGCCCTCCAAAGCAGCTCTATCAGTCTATCGCTAAACCCCAAC...C 647
192 LeSerPheAlaArgAsnArgAlaAspLeuAsnGlyAlaIleLysGly 208
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
648 TAACCTTTGCCAAAACAAAGCAAGCAAGAGGGGGTCCCTCTATTCC 697
209 SerAsnLeuIleCysSerGlyAsn...ValAsnProLeuPhePheThrGlu 224
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
698 ACGGAGGAGATTACAATTAACTATGCTTAACTGACATCATTTTCGTA 747
224 yAsnSerAlaThrAsn...::: 229
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
748 AAATACCGCGGAGCAATAGCGGAGCATTTACAGGAAGCTACAGATT 797
229 229
798 TTATTAGCAGCAACAAAGCAATTAGCTTTATTAACAATAGTGACCCCA 847
230GlyGlyAlaIleCysCysIleSerAspLeuAsnThrSe 242
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
848 ACCTACCTACAGGCGGAGCACTTACTGTAAGTACT...ACATTCACG 891
242 rGluLysGlySerLeuSerLeuAlaCysAsnGlnIleuThrLeuPheAla 259
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
892 CCCCAAACAGCTTAACTATCATCAGACACAGCGGAGCACTGACTTTATAG 941
259 erAsnSerAlaLysGluLysGlyAlaIleTyrAlaLysHisMetVal 275
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
942 GAAATACAGCAATTACTAGTGGGGGCAATTATCTGACAACTAGATT 991
276 LeuArgTyrAsnGlyProValSerPheIleAsnAsnSer..... 288
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
992 CTTTCTTCGAGAGCACTACCTTTTAAAAACAACCTCGCTATCATCATC 1041
289 ...AlaLysIleGlyAlaIleAlaIleGlnSerGlySerLeu 304
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1042 TGCAGCTCCCTTAGAGAGAGCAATGCGATCTGCTGATCTGATCTTGA 1091
304 erIleLeuAlaGluGluGlySerValLeuPheGlnAsnAsnSer..... 318
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1092 GTCTTCGCTCTTGGTGAGACATCACTTTTGAAGGAACACAGTACTGC 1141
319 ...GlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu... 333
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1142 AAAGGAGCTTCTCGAGTCAGACCACTACAGAAATTCATTAACATCCG 1191
334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...I 349
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1192 AAACACCAATGCTAGATGTACAGCTGCGAGCTCTCAAGCAATACTA 1241
349 LeuPhePheAspProIleValGlnGluSerSerLysGluSerPro 365
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1242 TCTACTTATATGATCTATTAACA..... 1264
366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1265ACTAGCATCTGACAGCTCTCTCAGATGCCCTAACTTAATAG 1308
382 aserProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 399
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1309 TCCCTGACCTTGACAGGAATCTCGATATCAAGAACCAATGATTTTCG 1358
399 erGluArgLeuSerGluGluGluLysThr...ProAspAsnLeuThrSer 414
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1359 GAGAGAGCTCTCGAAGCAAGAGCTGCAGAAAGCTGATATCTCAAAATCT 1408

415 GlnLeuGlnIleProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1409 ACAATTCAGCAACCTTAACACCTTGGCGGAGGCGCAACTCTCTTAATATC 1458
431 parAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1459 AGGAGTCACTTGTAGTCTTTCGCAATCCCGGCTCTAC 1508
448 euIleMetGluAlaGlyThrSerLeuLysThrSerAspLeuLysLeu 464
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1509 TCTCTATGATGACAGGACACATTAAGAACCGCTGATGGATCATATTC 1558
465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGluLysSerValTh 481
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1559 AATAATCTCTTCATCATGTAGATTCCTTAAGACAAAGAGGCGCAC 1608
481 rIleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1609 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTATCTGATCGC 1655
498GlyAspLysAsnPheGlyArgLysValGluLeu 509
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1656 TCTCTCTTGTAGATCTCTCGAAATGCTTACGAAAGATGC..... 1696
510 SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLys 523
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1697TCTGGAAATACCCCTCAAGTCTTCTGTCTGCTCACTCTTACGCG 1740
523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer.... 537
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1741 TGACGACCCCGCAATATTCATCAGACAGAC.....TTACTCTGATC 1784
538HisPheGlyTyrGlnGlyAspTyrPhe 547
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1785 CCGTAAAGAAAAATCTTATCCATTTGGGATACCAAGGAATGGGCAATTA 1834
548 SerTrpLysAspSerAspGluGlyHisSerLeuIleAla.....AsnTr 562
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1835 TCTTGCAAGAGATATCGCGCAATATCCAAAGAGGAGCTTACTCTG 1884
562 pThrProLysAsnTyrValProHisProGluArgGlnSerThrLeuVal 579
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1885 GACAAACAGAGATACATCCGAATCCGAGCGTGGTGAACCTTAGTTG 1934
579 IaAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValGlnSerMet 595
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1935 CTACACCCCTATGGGATCTTTGTTGATGTGCGCTCATACACAGCTT 1984
596 IleAsnThrIleAlaHisGlyAlaTyrLeuPheGlyThrTrpGlySe 612
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
1985 GTACCCACTAAAGTACGCCAATCTCAAGAACTCGCGCATCTGGTGTA 2034
612 rAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIle 629
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2035 AGGGATCTCGAACTTCTC...CATAAAGATGACAGGAATTAATA 2078
629 sPasnTrpHisThrArgSerLeuGlyTyrLeuPheGlyIleSerThrHis 645
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2079 AAGCTTTTCGCCACATAAGTCAGGTTATGTTGAGGAGCACTACACA 2128
646 SerLeuAspHisSerPheCysLeuAlaIleGlnGlnLeuGlyLys 662
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2129 TTAGCTTCTGAATATCTTATACCTGACGCTTTCGCAATTAATTCGCAA 2178
662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyrIleAla 679
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2179 AGATGAGATCACTTATTAATAAAAGAGCTTCTCTGATGAGCTT 2228
679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
2229 CTCTCATCTCCAGCATCTAGCAGACCTTCTCTTCCAAAGCTTGTACGC 2278
691IleSerAlaGlnAl 695

105091 AAATACCGGCGGCAATGGCGAGCCATTACAGGAGCTAGCACTT 105042
 229
 105041 TTATTAGCAGCAACAAACCAATTAGCTTTATTAACAAATAGTGTACCCCA 104992
 230glyAlaIleCysCysIleSerAspLeuAsnThrSe 242
 104991 ACCTCAGCTACAGGGGAGGACATTTACTGTAAGTACT.....ACATTCAGC 104948
 242 rGluLysIleSerLeuSerLeuAlaCysAsnGlnGluThrLeuPheAla 259
 104947 CCCCACCAACCTTTACTTACTTACAGACACCGGAGACTGAACCTTTATAG 104898
 259 eAsnSerAlaLysGluLysGlyAlaIleTyrAlaLysIleMetVal 275
 104897 GAATACAGCAATTTACTAGTGTGGGCGGATTTTACTGACAACTTAGTT 104848
 276 LeuAlaGlyTyrAsnGlyProValSerPheIleAsnAsnSer..... 288
 104847 CTTCCTTCGAGAGCACTACCGCTTTTAAACCAACCTCTGCATAGATAC 104798
 289 ...AlaLysIleGlyAlaIleAlaIleGlnSerGlySerLeu 304
 104797 TGCACCTCTCCAGAGGAGCAATGCGATGCTGACCTGATCTTGA 104748
 304 eTlleLeuAlaGlyGluLysSerValLeuPheGlnAsnAsnSer..... 318
 104747 GTCTTGGCGCTCTGGTGAGACATCACTTTTGAAGAAACACATAGTC 104698
 319 ...GlnAlaGlyThrSerAspGlnGlyLeuValArgAsnAlaIleTyrLeu 333
 104697 AAGACGCTCTTCGAGTACAGACCACTACAGAAATCTTATTAACATCGG 104648
 334 .GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAsp...I 349
 104647 AACACCAATCTAGATTGATGATGACCTCGAGCCCTCAGAGCACTACTA 104598
 349 lLeuPhePheAspProIleValGlnGlnSerSerGlySerPro 365
 104597 TCTACTTCTATGATCTTATACA..... 104575
 366 LeuProSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAl 382
 104574ACTACATCAGTCAGACGCTCAGATGCTTAACCTTAATAGC 104531
 382 aSerProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSer 399
 104530 TCCTGACCTTCAGGGAATCTGCTATCAAGAAACCATCGATTTTCTG 104481
 399 eArgLysLeuSerGluGluGluLysThr...ProAspAsnLeuThrSer 414
 104480 GAGAGACCTCTCGAGACAGAACTGCAGAACTGATATATCTCAAACT 104431
 415 GlnLeuGlnIleProIleGluLeuLysSerGlyArgLeuValLeuLysAs 431
 104430 ACAATTCAGCAACCTTAACCTTCGCGAGGAGCAACTCTCTTAATATC 104381
 431 pArgAlaValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeu 448
 104380 AGGAGCTCACTAGTCTAGTCTTTCCTTCGCAAACTCCGGGCTTACCC 104331
 448 eLleMetGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeu 464
 104330 TCCTATGATGATGAGGAGGACCATTAAGAACCGCTGATGGATCTCAATC 104281
 465 AlaThrLeuSerIleProLeuHisSerLeuAspThrGlyLysSerValTh 481
 104280 AATAATCTTGTTCATCATCTAGATCTTAAAGAGACCAAGAACGCTAC 104231
 481 rLleHisAlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSer 497

104230 GCTAAAGCAACACAGCAAGT...CAGACAGTCACTTATCTGGATCG 104184
 498GlyAspGluAsnPheTyrGluAsnValGluLeu 509
 104183 TCTCTCTGTAGATCCCTTCTGAAATGCTCTACGAAGATCTC..... 104143
 510 SerLysGluGlnAsnAsn.....IleProLeuLeuThrLeuSerLys 523
 104142CTTCGAATAAACCTCAAGTCTTCTGTGTCTACTCTTACTTC 104099
 523 sGluGln...SerHisLeuHisLeuProAspGlyAsnLeuSerSer... 537
 104098 TGACGACCCCGGAATATATCAATCAGAC.....TTCACCTCTGATC 104055
 538HisPheGlyTyrGlnGlyAspThrPhe 547
 104054 CCGTAGAAAAAATCCATTCATTCGATGGGATACCAAGGAATGGGCACTA 104005
 548 SerTrpLysAspSerAspGluGlnHisSerLeuIleAla.....AsnTr 562
 104004 TCTTCGCAAGAGAACTGCGACATTAATCCAAAGCAGCAGCTTACCTG 103955
 562 pThrProLysAsnTyrValProHisProGluArgGlnSerThrLeuVal 579
 103954 GACAAAAACAGATACAAATCCGAATCTGAGCGTGTGGAACTTAGTTG 103905
 579 lAsnThrLeuTrpAsnThrTyrSerAspMetGlnAlaValGlnSerMet 595
 103904 CTACACCCCTATGGGAGATCTTGTGTGATGTGCGCTCATCAACAGCTT 103855
 596 lLeasnThrIleAlaHisGlyAlaTyrLeuPheGlyThrTrpGlySe 612
 103854 CTAGCCACTAAGTACGCCAATCCAAAGAACTCGCGCATCTGTGTA 103805
 612 rAlaValSerAsnLeuPheTyrAlaHisAspSerSerGlyLysProIle 629
 103804 AGGGATCTCGAATCTTCTC.....CATAAAGATAGCAGGAATATA 103761
 629 sPasnTrpHisIleArgSerLeuGlyTyrLeuPheGlyIleSerThrHis 645
 103760 AAGCTTTGCGCACATATAGTCACGCTTATGTTGAGAGCCACCAACA 103711
 646 SerLeuAspAspHisSerPheCysLeuAlaAlaGlyLeuLeuGlyLys 662
 103710 TTAACCTCTGAAATCTTATCACTGACCTTCGCCAATTAATTCGGGA 103661
 662 sSerSerAspSerPheIleThrSerThrGluThrThrSerTyrIleAla 679
 103660 AGATAGAGATCACTTATTAATAATAATAGAGCTTGTGCTATGACAGCT 103611
 679 hrValGlnAlaGln...LeuAlaThr.....SerLeuMetLys 690
 103610 CTGTCCATCTCCAGCATCTAGCAGACTTGTCTCTCCAGCTTGTACGC 103561
 691IleSerAlaGlnAl 695
 103560 TACCTCTCGATCTGAAGAGTACAGAGCTGTCTCTTGTATGCTCAAG 103511
 695 acysTyrAsnGluSerIleHisGluLeuLysThrLysTyrArgSerPhe 712
 103510 CAGCTATATCTATAGTAAATACTATGAAAACTATTACACCCCAAGCAC 103461
 712 eLysGluGlyPheGlySerTrpHisSerValAlaValSerGlyGluVal 728
 103460 CAAGC...GGAAGAGAGCTCGTATATATGACGGTGGCGCTCGAAGCT 103414
 729 CysAlaSerIleProIleValSer...AsnGlySerGlyLeuPheSerSe 744
 103413 GCGAGCTCTCCATCACACATCGCTTAAAGCATGAGCGCTCTTCCACAG 103364
 744 rPheSerIlePheSerLysLeuGlnGlyPheSerGlyThrGlnAspGlyP 761
 103363 GTATTTCTTTCATCAAGTAGAAGCTTCTGATACACCAAGATAGCT 103314

[illegible]

472013 ATGAGAGCCTCTTCTCTCTCTTGTGCTCATGATCACAAGCAGTCC 472062
67 roleuTyValleuGlyAsnSerTyCysTrpPheValSerLysLeuHis 83
|||||
472063 CTTTGATGATGTAAGAAATTCCTACTGTTGCTGATCTAACTGCAT 472112
84 lIethAsPProIySGuAlaLeuPheLysGluLysGlyAspLeuSerlI 100
|||||
472113 ATCAGGAGCCCAAGAGAGCCTCTTTTAAAGAAAAAGAGATCTTCCAT 472162
100 eGIAsnPheArGpHeuSerPheThrAspCysSerSerLysGluSers 117
|||||
472163 TCAGAACTTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAAGGAAACCT 472212
117 erProSerlIeIleHISGlnLysAsnGlyGlnLeuSerLeuArgAsnAsn 133
|||||
472213 CTCCTCTATATTATTCATCAAAAGATGGTCAGTATCTCTGCCCAATAT 472262
134 GlySerMetSerPheCysArgAsnHisAlaGluClySerClyAlaIle 150
|||||
472263 GGTAGCATGAGTTCTGTGCAATCATGCGAAGCCTGTGAGAGCCAT 472312
150 eSerAlaAspAlaPheSerLeuGlnHisAsnTyrlLeuPheThrAlaPheG 167
|||||
472313 CTTCTGGCATCCCTTTTCTCTCTACAGCACAATCTTTTTCACACCTTTTC 472362
167 lUgluAsnSerSerLysGlyAsnGlyAlaIleGlnAlaGlnThrPhe 183
|||||
472363 AAGAGATTCCTCTTAAGGAATGGCGAGCCATTGAGCTCAACCTTC 472412
184 SerLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAs 200
|||||
472413 TCTTTTCTAGAAATGTGCGCTATTCTTCTGCCCCATATGCTGGGGA 472462
200 pLeuAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerGlyAsnTy 217
|||||
472463 TTTAAATCGCGCGCCCTATTTCCTGAGCAATCTTATTTGTCAGGAGAAAG 472512
217 aIAsnPProIeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaIle 233
|||||
472513 TAAACCTCTCTTTTCTACAGCAACTCCGCCAGAAAGAGAGCGCTATT 472562
234 CysCysIleSerAspLeuAsnThrSerGluLysGlySerLeuSerLeuAl 250
|||||
472563 TGTGTGATAGGCATCTAAACACCTCAGAAAAAGGCTCTCTCTCTGCG 472612
250 acYAsnGlnGluThrLeuPheAlaSerAsnSerAlaLysGluLysGlyG 267
|||||
472613 TTCTAACCAAGAAACCTATTTCGAAGCAATTCGCTAAAGAAAAAGCGC 472662
267 lYAlaIleTyAlaLysHisMetValLeuArgTyArgAsnGlyProValSer 283
|||||
472663 GGGCTTTTATGCGCAAGCACATGCTATTCGCTTATACGCTCTGTTTGC 472712
284 pheIleAsnAsnSerAlaLysIleGlyAlaIleAlaIleGlnSerC 300
|||||
472713 TTTCATTAAACAGCCCTAAATAGGTGAGCTATCGCATCCAGTCCGG 472762
300 yGlySerLeuSerIleLeuAlaGlyGluLysSerValLeuPheGlnAsnA 317
|||||
472763 ACGGACTCTCTATCTCTGACGAGTGAAGATCTGTTCTTCACAAGAA 472812
317 snSerGlnArGThrSerAspGlnGlyLeuValArgAsnAlaIleTyrlLeu 333
|||||
472813 ACTCCCAAGCAACCTCCAGCAAGGCTGTAAGAAAGCCCATCTTACTTA 472862
334 GluLysAspAlaIleLeuSerSerLeuGluAlaArgAsnGlyAspIleLe 350
|||||
472863 GAGAAAGATGCGATCTTCTTCTTAGAAGCTCGCAAGAGATATCT 472912
350 ubPhePheAspProIleValGlnGluSerSerSerLysGluSerProLeuP 367
|||||
472913 TTTCTTTGATCTCTATTCTACAAAGAAAGTAGCAGCAAGAAATCCCTTTC 472962

367 roSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSer 383
|||||
472963 CCTCCCTTTTGGCAAGCAGGTGACTTCTCCACCCAGCACCGGATCT 473012
384 ProLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerG 400
|||||
473013 CCTTGATGATTATCAGCAAGTCGCAAAACCGTTCAGTATTTTCTCGAGCA 473062
400 uArgLeuSerGluGluGluLysThrProAspAsnLeuThrSerGlnLeuG 417
|||||
473063 ACCTCTTTCGAGAGCAAAAAAACCCTGATACCTGATCCACTCCCACTAC 473112
417 lInGlnProIleGlnLeuLysSerGlyArgLeuValleuLysAspArgAla 433
|||||
473113 AGCAGCCTTACGACGTAATCCGAGCGCTGTGTTTAAAGATCCGCT 473162
434 ValLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuIleMe 450
|||||
473163 GTCCTTCCCGCGCTCTCTCTCTCAGATCCCAAGCTCTCTCTCATTTAT 473212
450 lGluAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThrL 467
|||||
473213 CGAACCGGAGACTTCTTAAAAAATCTCTGATTTGAAGTTACT ACCC 473261
467 euSerlIeProLeuHisSerLeuAspThrGluLysSerValThrIleHis 483
|||||
473262 TAAGTTTCCCTTCACTTCTTAGATATCGAAAAAGCCATCACTATCCAC 473311
484 AlaProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspC 500
|||||
473312 GCCCTTAATCTTTCATCCAAAAAGATCTTCTCTCAACTCGAGATGA 473361
500 uAsnPheTyrlGluAsnValGluLeuLeuSerLysGluGlnAsnAsnIleP 517
|||||
473362 GAATTTTATGAAAAATGTAAGCTTCTCACTTAAGAGCAAAACAATATTC 473411
517 roleuLeuThrLeuSerLysGluGlnSerHisLeuHisLeuProAspGly 533
|||||
473412 CTTCTCTTACTCTCTCTTAAGAGCAATCTCATTTTACATTTCTCTGATGG 473461
534 AsnLeuSerSerHisPheGlyTyrlGlnGlyAspTrpThrPheSerTyrl 550
|||||
473462 AACCTCTCTTTCACATTGATGATCAAGAGATGAGACTTTTCTTGGA 473511
550 sAspSerAspGluGlyHisSerLeuIleAlaAsnTrpThrProLysAsnT 567
|||||
473512 AGATTCTGAWGAAGGCAATTCCTGATTCCTAATGGAAGCGCTAAAAACT 473561
567 TyrlValProHisProGluArgGlnSerThrLeuValAlaAsnThrLeuTrp 583
|||||
473562 ATGTGCTTATCCAGACGTCATCTACACTCGTTGCGAACACTCTTGG 473611
584 AsnThrTyrlSerAspMetGlnAlaValGlnSerMetIleAsnThrIleAl 600
|||||
473612 AACACCTATTCGATATGCAAGCTGTGAGTGATGATTAACAAACAGC 473661
600 aHisGlyGlyAlaTyrlLeuPheGlyThrTrpGlySerAlaValSerAsnL 617
|||||
473662 GCACGAGAGAGCCCTATCTATTGGAACGAGCACTGCTGCTTCTAAT 473711
617 euPheTyrlAlaHisAspSerSerGlyLysProIleAspAsnTrpHisHis 633
|||||
473712 TATTCATGTTTCAGCAGAGCTCTGGAAAACTATGATTAATTTGGCATAT 473761
634 ArgSerLeuGlyTyrlLeuPheGlyIleSerTrpHisSerLeuAspAspH 650
|||||
473762 ACAAGCCTTGCTTACTATTCGCTATCACTACTCACTGATTAAGATGACCA 473811
650 sSerPheCysLeuAlaAlaGlyGlnLeuLeuGlyLysSerSerAspSerP 667
|||||
473812 TTTCTTTGCTTGGCTTCGAGGACAAATTTACTCGGAAATTCGTCATTCCT 473861

Align seg 1/1 to: A64751 from: 1 to: 2643

```
1 MetArgProAspHisMetAsnPhenCysCysLeuGlyAlaAlaLeuLeu 17
|||||
7 ATGGAGCTGATCATATGACTTCTGTCTATGTCGCTATTTTCTC 56
|||||
17 rSerThrAlaValLeuPheGlyGlnAspProLeuGlyGlnThrAlaLeu 34
|||||
57 ATCCACACGGCTCCTTTGGCCAGGATCCCTTAGTGAAACCGCCCTCC 106
|||||
34 eutThrLysAsnProAsnHisValAlaLysThrPhePheGlnAspCysThr 50
|||||
107 TCACATAAAATCTTAATCATCTCTCTGTACATTTTGTAGACACTGTACC 156
|||||
51 MetGlnSerLeuPheProAlaLeuGlyAlaHisAlaSerGlnAspAsp 67
|||||
157 ATGGAGAGCTCTTCTCTCTCTCTCTCTCATGATCACAAGAGATGCC 206
|||||
67 oLeuTyValLeuGlyAsnSerTyGlyTrpPheValSerLysLeuHis 84
|||||
207 TTTGTATGACTTGGAAATTCCTACTGTTGGTTCGTATCTAAACTCCATA 256
|||||
84 leThrAspProLysGlnAlaLeuPheLysGlyLysGlyAspLeuSer 100
|||||
257 TCACGACGCCCAAGAGGCTCTTTTAAAGAAAAGAGATCTTTTCCATT 306
|||||
101 GlnAsnPheArgPheLeuSerPheThrAspCysSerSerLysGlySer 117
|||||
307 CAAAACCTTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 356
|||||
117 rProSerIleIleHisGlnLysAsnGlyGlnLeuSerLeuArgAsnAsn 134
|||||
357 TCCTTCTATATTCATCAAAAGATGGTCAGTATTCCTTGGCCATATATG 406
|||||
134 LysSerMetSerPheCysArgAsnHisAlaGlyLysGlyValAlaIle 150
|||||
407 GTACCATAGTTCTGTCGAATCATGCTGAAGGCTCTGACGACCCTATC 456
|||||
151 SerAlaAspAlaPheSerLeuGlnHisAsnTyLeuPheThrAlaPheG 167
|||||
457 TCTGGGATGCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 506
|||||
167 uGlnAsnSerSerLysGlyAsnGlyAlaAlaGlnAlaGlnThrPhe 184
|||||
507 AGAGATATCTTCTTAAAGCAATGGCGGACCATTCAGGCTCAAAACCT 556
|||||
184 eLeuSerArgAsnValSerProIleSerPheAlaArgAsnArgAlaAsp 200
|||||
557 CTTTATCTAGAAATGTCGCCCTATTTCTTCCGCCGTAATCTCCGCAT 606
|||||
201 LeuAsnGlyAlaAlaIleCysCysSerAsnLeuIleCysSerGlyAsn 217
|||||
607 TTTAAATGGCGGCGCTATTTGCTGTAGTATCTTATTGTTCAAGGATGT 656
|||||
217 AsnProLeuPhePheThrGlyAsnSerAlaThrAsnGlyAlaAlaIle 234
|||||
657 AAACCCCTCTCTTTCTACTGGAACCTCCGCCAGATAGAGGCGCTATTT 706
|||||
234 ySCysIleSerAspLeuAsnThrSerGlnLysGlySerLeuSerLeuAl 250
|||||
707 GTTGATTCGCGATCTAAACACCTCAGAAAAGGCTCTCTCTCTGTGCT 756
|||||
251 CysAsnGlnGlyThrLeuPheAlaSerAsnSerAlaLysGlyLysGly 267
|||||
757 TGTAAACCAAGAAACGCTATTTCAGACCAATTCGTCTAAAGAAAAGCG 806
|||||
267 yAlaAlaIleTyAlaLysHisMetValLeuArgTyAsnGlyProValSer 284
|||||
807 GCGCTATTTATGCCAAGCATGATGCTATGCTATTAACGGTCTCTGTTCT 856
|||||
284 heIleAsnSerSerAlaLysIleGlyAlaAlaIleAlaIleGlnSerGly 300
|||||
857 TCATTTAACACAGCGCTAAATAAGTGGACCTATCCCATTCAGTCCGGA 906
|||||

301 GlySerLeuSerIleLeuAlaGlyGlnGlySerValLeuPheGlnAsn 317
|||||
907 GGGAGTCTCTCTATCTCTGCAAGGTGAAGATCTGTCTGTCCAGAAATA 956
|||||
317 nSerGlnArgThrSerAspGlnGlyLeuValArgAsnAlaIleTyLeu 334
|||||
957 CTCCCAAGCAGCTCCGACCAAGGCTAGTAAGAAACCCATCTACATTA 1006
|||||
334 LysAspAlaIleLeuSerSerLeuGlnAlaArgAsnGlyAspIleLeu 350
|||||
1007 AGAAAGATCCCATTTCTTCTCTTAGAAGCTCCACAGCGATATTTCTT 1056
|||||
351 PhePheAspProIleValGlnGlySerSerSerLysGlySerProLeu 367
|||||
1057 TTCTTTGATCTATGTTACAGAAAGTAGCAGAAAGATGCGCTCTCC 1106
|||||
367 oSerSerLeuGlnAlaSerValThrSerProThrProAlaThrAlaSer 384
|||||
1107 CTCCCTTTGCAAGCAGCGTACCTCTCCACCCACCGCATCTC 1156
|||||
384 rLeuValIleGlnThrSerAlaAsnArgSerValIlePheSerSerGly 400
|||||
1157 CTTTAGTTATTCACACAGTCAAGTCAAAACCTTCAGTATTTTCTCGAC 1206
|||||
401 ArgLeuSerGlnGlyLysLysThrProAspAsnLeuThrSerGlnLeu 417
|||||
1207 CGTCTTTCTGAAGAAAGAAACCTCTGATTAACCTCACTTCCAACTACA 1256
|||||
417 nGlnProIleGlnLeuLysSerGlyArgLeuValLeuLysAspArgAla 434
|||||
1257 GCAGCCTATTCACACGAAATCCGAGCGTTAGTTTAAATAATCCCGCTG 1306
|||||
434 AlLeuSerAlaProSerLeuSerGlnAspProGlnAlaLeuLeuMet 450
|||||
1307 TCCTTTCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1356
|||||
451 GlnAlaGlyThrSerLeuLysThrSerSerAspLeuLysLeuAlaThr 467
|||||
1357 GAAGCGGAACCTCTTAAAACTCTCTGATTTGAAGTTAGTACACCT 1406
|||||
467 uSerIleProLeuHisSerLeuAspThrGlnLysSerValThrIleHis 484
|||||
1407 AAGTATTCCTCTCATTCCTTAGATAGTGAATAAAACCTATATCCACG 1456
|||||
484 LeProAsnLeuSerIleGlnLysIlePheLeuSerAsnSerGlyAspG 500
|||||
1457 CCCCTAATCTTTCTATCCAAAAGATCTTCTCTCTACTCTGAGATGAG 1506
|||||
501 AsnPheTyGlnAsnValGlnLeuLeuSerLysGlnLysAsnAsnIle 517
|||||
1507 AATTTTATGAATAATGTAGAGCTCTCAGTAAAGCAAAACAATATTC 1556
|||||
517 oLeuLeuThrLeuSerLysGlnGlnSerHisLeuHisLeuProAspGly 534
|||||
1557 TCTCTTACTCTCCCTTAAAGACCAATCTCATTTACATCTCTCTGATG 1606
|||||
534 snLeuSerSerHisPheGlyTyGlnGlyAspTrpThrPheSerTrpLys 550
|||||
1607 ACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1656
|||||
551 AspSerAspGlnGlyHisSerLeuIleAlaAsnTrpThrProLysAsn 567
|||||
1657 GATTCGTGAAGAGGCGATCTCTGATTTGCTAATTTGAGAGCCCTAAAC 1706
|||||
567 rValProHisProGlnArgGlnSerThrLeuValAlaAsnThrLeuTrp 584
|||||
1707 TGTGCTCATTCAGAACGTCATTAACACTCTTTCGAAACACTCTTTTGA 1756
|||||
584 snThrTySerAspMetGlnAlaValGlnSerMetIleAsnThrIleAla 600
|||||
1757 ACACCATTTCCGATATGCAAGCTGTGACACTGCATGATTAATTAACAAG 1806
|||||
```


QY	295	IAIOGSGSLIAGCGSLFPMNSQRTSDOGLV--RNAIYLEKALLISLEARNGD-ILF	351
Db	344	IIInGgSLIqIItGgqgdlfEqnqvtvltlapatktkrnIlnhEstakwCtGlaaSqnaIyF	4030
QY	352	FDPIVOESSKESPLPSSLAQSVSPPTATSPVITQTSANR-----SVIFSSSEALSEEK	4070
Db	404	ydpI-----tctndgaasnIrlInEsaqrkIsgsIysIsgEIsaaE	4450
QY	408	TPDNTSLOQOPIELKSGRLVKKRAVLAPSLSDQPOALLMEAGTSLKTS--DLKLAT	4660
Db	446	IaenItfcrngpvtIvegsIeIkqgvltIlgIsqEpeStIIIdIgtIsIqastEdIvItn	5050
QY	467	LSIPHSLDTEKSVTIAH-----PNIASIQKIFLMSGCPENFEVENELLSKONNIRPLT--	5200
Db	506	ssInadtlLygnpIlnIaEaankIltIglalIvnaqalIyenIIdqsdqdsIsvkIsPg	5650
QY	521	-----LSKEQSHLHPDGNLSHFYGOGDWTFSWKDSDEGHSLIAN--WTPKNVPHPER	5730
Db	566	aggtIttgdaSqkIlevaSprtPhygyghnmvqyIpgtGtqpsatIlewrIttyIInper	6250
QY	574	QSTLVANTLNMITYSDMQAVQSMINTIAHGAYLFETWGSVAVSNLEFVHDSGKPIDNMHH	6330
Db	626	qgflvpmIswsIvdqrfIgeImwssqIlceqergvvaqIdan-fIhrckIInE--hgyrth	6820
QY	634	RLSLGLFQISIHSLDDHDFCLAAQOLGKSSDSFITSFTEHTSYIAVY-----	6800
Db	683	sgvgyIvvgthafsdalInaIcqlIfrdkdIvvasKnhgIysgvvIledIlefspqg	7420
QY	681	-----QAOLATSLMKIASQACYNESIHELTKRYRSFKEGFGSMHSAVAGECASIPR	7340
Db	743	fytIdssEaecnqvtIdmgIysghrmdmkItkyty-pegqswandvfyIetgaltty	8010
QY	735	VSNQGLSSFSIStKLOGFSGTQDDGEESGELIRSSASFRNISLPITTEPKKSQKT	7940
Db	802	ypnsfItIdySpIrlIqClyahgedIkeIgevrhItsgdIlnIavpIgvkIerIsdcK	8610
QY	795	RTYVYFLGAYIQD-LKRGVESCPVVLKNAVSDMAPANLDSRAYMPRIITNQALH-RLQ	8520
Db	862	rgsyglItIayppdvIrkrpkE--faIlaSgIwslbhqnmIsrgIqlrIghnclnpgIe	9190
QY	853	TLLNWSCVLRGOSHYSYLDLITTRF	8780
Db	920	vfhgaIeIrgstrnylnIhgkyrf	9450
RESULT 23			
ID	Y96274	standard; Protein; 973 AA.	
AC	Y96274:		
DT	10-NOV-2000	(first entry)	
XX			
DE		Chlamydia POMP91B precursor protein.	
XX			
KW		Chlamydia; POMP91B; respiratory tract disease; infection;	
KM		bronchitis; sinusitis; pneumonia; atherosclerosis; asthma; vaccine;	
KW		immunogen.	
XX			
OS		Chlamydia pneumoniae.	
XX			
PN	WO2000262319-A2.		
XX			
PD	11-MAY-2000.		
XX			
PF	02-NOV-1999;	99WO-CB03622.	
XX			
PR	02-NOV-1998;	98US-010650.	
XX			
PR	07-MAY-1999;	99US-013071.	
XX			
PR	29-OCT-1999;	99US-0430723.	
XX			
PA	(CONN-)	CONNAUGHT LAB LTD.	

PI Murdin AD, Oomen RP, Dunn PL;
DR WPI: 2000-365571/31.
XX N-PSDB; A27342.
XX
XX Novel Chlamydia POM91B precursor protein antigen, used for vaccination
XX and protection against Chlamydia infection
XX
XX Claim 6; Fig 1: 97pp; English.

The present sequence is the chlamydia pneumoniae POM91B precursor protein. Infection by Chlamydia can result in respiratory tract diseases such as bronchitis, sinusitis and pneumonia, asthma and atherosclerosis. The gene, protein and antibodies can be used as immunogens to induce an immune reaction in humans which has the effect of vaccinating the person. They can also be used to diagnose and treat those infected with the parasite.

SQ Sequence 973 AA:

Query Match 17.9%; Score 810.5; DB 21; Length 973;
Best Local Similarity 26.3%; Pred No. 3e-59;
Matches 251; Conservative 158; Mismatches 405; Indels 139; Gaps 25.

37 NNHVAVCTFEEDCTMESLFPALCAHSA-----QDDPLYLVGNSYCMFYSKHLITDPEKAL 91
:
Db dbrgltclctfsglylandlnalnsrlssscfnragalqlgygfslntlsadgaai 108
49
QY 92 FK-----EKGLSTIONRFLSFTDCSSKESSPS-----IHKQ---NCQLSLRNNGSKS 137
:
Db 109 ssvitqnpelcplcsifsfsgfmfdncestlstsdasanyphasaiayatpmflfnndsll 168
138 FCERNHAEGSGGAISADAFSLQHNYLFETAPEENSSKGNGAIG-AQTFSIRNVSPISFAR 196
:
Db 169 fgmtnsaagaaiairysltstlenticskslifngngslnggalcgasaainlmhsapylfst 228
197 NRADLNGAICCS--NLICSGVNDLPFTGNSATNCGAICTSIDNTSEKG----- 245
:
Db 229 natlglygaalytggsmultsgnlsqvlfvnnssrgaiayanvnlfnmsdltfgnota 288
246 -----SLSLACNQETIFASNAKE 264
288 spqmshlpptcpptpavtrpllgygaafctcpatprrptqvavtltsgensvtflenlse 348
265 KGGAIYAKHMVRYRNGPVSFINNASKIGGAIAIQSGSLSLAGEGSVLFQNNSORTSDQ 324
:
Db 349 gggaglygkklisidsnkstfligntcggkaiaipesgeislanaagdillfknlsltsgt 408
325 GLVRNAIYLEKNAILSSLERANG-DILFFDPDIVOGSSSKESPPLPSSLOASVTSPTPATA 383
:
Db 409 p-trnsinhgkakfatlgntgtylfydpfltsddlsaas-----aaa 451
384 PLYIVTOSAIR-----SVIFSSEKL-SSEEKRPDMLTSQLOPIELKSRLYLKORAVLS 436
:
Db 452 lvvvnpksadagaaygvtifsgeltateaatpaatstlnklelegaltialmgatln 511
437 APSLSODPOLLIIMAGTSALKRSSD-----LKLATISIPHSLDTEKSVTHAPN--- 486
:
Db 512 vnhftqdeksvviimdagttlatctngannotdgatclinklvindslidgtkaaavnygstng 571
487 -LSIQTFILSNSGDENPFENVELLSEKQNNIPLLTLSEKQSHLHPD-----GNLSHF 539
:
Db 572 alitglctlglyvnsgdcdbnmfknldlqgvplilalkatsnvtctldfaelgnvgqspdy 631
540 GVQGGMWTFSMKSDSGHSLIAMWPKNVPVPHEROSTLYANLMTMYSMQMVQSMLINTI 599
:
Db 632 gygagwefetfidtc--hvctgmwkktcyrlpripertlaplpslwianvdlfrvsaga--sa 687
600 AHG----GAYLEGTWGSAAVSNLFYA-HDSGKPIDNMHHRSLGYLEFGISTHSLDHSFCL 654
688 adgedvpapkqslst---qlntffnahnhtcdar---svltmqqvylinvrtlrldaalsl 741


```
OY 226 SAT-----NGAICISLNLNTEKSGSLACNOETLPFASNSAKEGCAIYAKKHVLRNGP 281
      |||||
      : : : : :
Db 252 svlaetaggal-----sgdadvtlsqngsvtfsqngavangalYakKklllae999 303
OY 282 VS---FINN-----SAKIGAIATQSGSLILAGEGSLVLFQNNNS-QRTSOGGLVRNAYL 333
      |||||
      : : : : :
Db 304 ggnpfsmivgttaaggagstilaagecslfseagqylnagatvltqtktrnsidi 363
OY 334 EKDALISLEARNG-DLFFDPPIVQESSKESPLPSLSLQASVTSPPATASPLVIQTSAN 392
      |||||
      : : : : :
Db 364 gsgqkthelraishgslffldpplantaadstldlnkhdag-----stdys 412
OY 393 RSVIFSSERLSEER-KTPDNLTSQLAQPIELKSGRLVLRDAVLASVLSQDQALLIME 451
      |||||
      : : : : :
Db 413 gsvlfsgklsedeakvadnltsllkqpvltagnvlkqgvltldtkgftqagssvimd 472
OY 452 AGTSLKTS--DLKATLSIPLHSLDTEKSVTHAP-----NLSIQIFLNSGDENFYE 504
      |||||
      : : : : :
Db 473 agtlkasteavltlgislpvdsigegkvviaasaknvalsgp11ldng--naye 530
OY 505 NVELLSKEON-----NIPULTSKESQSHLHPDGNLSHIFGQDPMFSMKD---SDEGH 556
      |||||
      : : : : :
Db 531 nht-1gkltqdfstfvglaalgtatctdvpaavpvauplhvgyqglwmlwddtasipk 589
OY 557 SLIAMPKPYVPHPERQSHLVANTLWNTVSDQAVQSMINTIAHGAFLGCTWGAASN 616
      |||||
      : : : : :
Db 590 lallawenlglpmpexgplvpnsllwgsldqalqivlarsalltcsdgtfwaagvan 649
OY 617 LFAHDSGCRPIDMMHRSGLYLFISTHSDHSFCLAQGLLGRKSSDSFITSTFTSY 676
      |||||
      : : : : :
Db 650 -fldkdkkgek-ikyrhksqgyaigaaqtcsenlifaftcgtfsgskdtfvaknhdty 707
OY 677 IATVQVQVQVATSL-----MKISAQCYNESIHEKTYRSPFSGKFG 717
      |||||
      : : : : :
Db 708 agatylhileecsfllgclldktpgswshkplvlegqlayshvndklkylay-pek 766
OY 718 SWSVAVSGEVCASIPVNSGSLFSSFSIFSKLQGFSGTQDCFEESGGEIRSFSSASF 777
      |||||
      : : : : :
Db 767 swgnafnmmlgaashsypelyhctdyapiklnlylqdsfexkgtegrstdsnlf 826
OY 778 NISLPITFEKSKQKRTTYFLGAIYQDLKRDVEGSPVLLKNVSMAPMNLDSRA 837
      |||||
      : : : : :
Db 827 nlslpivktekfcdndfslsydpdlindpckctalvisgawetwyanlarga 886
OY 838 YMFRLTQNALHRL-QTLNWSVLRGOSHSYSLDLGTTYRF 878
      |||||
      : : : : :
Db 887 lqvtaagshyalspmfevlqgfvleivrgrsrllynvdlggkqf 928

RESULT 16
W88429
ID W88429 standard. Protein: 914 AA.
XX
AC W88429:
XX
XX 26-APR-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed protein Omp5.
XX
KM Omp5: outer membrane protein 5; surface exposed protein; antigen;
XX infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS Chlamydia pneumoniae.
XX
PN W09858953-A2.
XX
PD 30-DEC-1998.
XX
PF 19-JUN-1998: 98WO-DK00266.
XX
PR 23-JUN-1997: 97DK-0000744.
XX
PA (BIRK/) BIRKELUND S.
```

```
PA (CHRI/) CHRISTIANSEN G.
XX Birkelund S, Christiansen G, Knudsen K, Madsen A:
PI Myglind P;
XX WPI; 1999-105610/09.
XX N-PSDB; X06828.
XX
PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins
XX
XX Disclosure: Page 78-80; 115pp; English.
XX
XX This polypeptide is described as a subsequence of the claimed
XX novel surface exposed protein Omp5 (see W88418) of Chlamydia
XX pneumoniae, a human respiratory pathogen. The invention provides
XX a new species specific test for identifying mammals (including
XX humans) infected with C. pneumoniae. The test comprises detecting
XX antibodies specific for surface exposed proteins Omp4-Omp5 (see
XX W88417-28) or detecting nucleic acid fragments encoding them (see
XX X06816-27), especially by PCR. The proteins are also used in the
XX diagnosis of C. pneumoniae infection in mammals. The nucleic acids
XX and proteins can also be used in the immunization of mammals, the
XX nucleic acids being particularly useful as DNA vaccines for
XX effecting in vivo expression of antigens. The vaccines may also
XX prevent atherosclerosis and bronchial asthma, which are possibly
XX associated with C. pneumoniae.
XX
XX Sequence 914 AA:
XX
XX
XX Query Match 18 7%: Score 846.5; DB 20; Length 914;
XX Best Local Similarity 30.6%: Pred. No. 2.4e-92;
XX Matches 277; Conservative 137; Mismatches 383; Indels 107; Gaps 30;
OY 12 CAALSTPAVLFGQDPL-GETALLTKPNHVVVC---PFEDCTQESTL--FPALCAHAS 63
      |||||
      : : : : :
Db 20 csvtlaateenlgsdsfsgstntgltvtpknttclgtldtltgltlgnlqdsaalckgc 79
OY 64 QD--DPLVYGNSTQWVSKLHTDPKEALPKEGDLSIONFRPLFTDCSSKE--SSPS 119
      |||||
      : : : : :
Db 80 sdtleslftagkygs--lslfnlksaeg-----aalsvtldnslstgtsfslflaaps 132
OY 120 II-----HKNGLSLRNNGSMSECRNHAEGSGAISADAFSLQHYLTAFEPENS 170
      |||||
      : : : : :
Db 133 svltcpsgkavkcgqdlctdmgcllftkqdyceenggaatstknslkstgtsftegnk 192
OY 171 SKG---NGAIOAQ-TPSLRNWSPISFARNRADLNGAI-CCSNLIGSGNVNPLFEFTGN 225
      |||||
      : : : : :
Db 193 ssatgkkggalcalgtvdltnlptlfnlnlaaagalaatngncltgnls-lyfsen 251
OY 226 SAT-----NGAICISLNLNTEKSGSLACNOETLPFASNSAKEGCAIYAKKHVLRN 279
      |||||
      : : : : :
Db 252 svlaetaggal-----sgdadvtlsqngsvtfsqngavangalYakKklllae999 303
OY 280 GPVSFINN-----SAKIGAIATQSGSLILAGEGSLVLFQNNNS-QRTSOGGLVRNAYL 333
      |||||
      : : : : :
Db 304 gsvlfsmivgttaaggagstilaagecslfseagqylnagatvltqtktrnsidi 363
OY 334 EKDALISLEARNG-DLFFDPPIVQESSKESPLPSLSLQASVTSPPATASPLVIQTSAN 392
      |||||
      : : : : :
Db 364 gsvkltnlraishgslffldpplantaadstldlnkhdag-----stdys 412
OY 393 RSVIFSSERLSEER-KTPDNLTSQLAQPIELKSGRLVLRDAVLASVLSQDQALLIME 451
      |||||
      : : : : :
Db 413 gsvlfsgklsedeakvadnltsllkqpvltagnvlkqgvltldtkgftqagssvimd 472
OY 452 AGTSLKTS--DLKATLSIPLHSLDTEKSVTHAP-----NLSIQIFLNSGDENFYE 504
      |||||
      : : : : :
Db 473 agtlkasteavltlgislpvdsigegkvviaasaknvalsgp11ldng--naye 530
```

in the diagnosis of *C. pneumoniae* infection in mammals. The nucleic acids and proteins can also be used in the immunization of mammals, the nucleic acids being particularly useful as DNA vaccines for effecting in vivo expression of antigens. The CC vaccines may also prevent atherosclerosis and bronchial asthma, which are possibly associated with *C. pneumoniae*.

Sequence 928 AA:

Query Match 18.8%; Score 857; DB 20; Length 928;
Best Local Similarity 30.2%; Pred. No. 3,2e-63;
Matches 261; Conservative 123; Mismatches 367; Indels 114; Gaps 24;

```

74 SYCMFVSKLHITDPKELFKKGLSTIONFRPLSTFCSSKSSPSIIHQKNGLSLRNN 133
Db 118 tllqfssistiaspgsiltgkavs-----cst-----gslkfdkn 154
QY 134 GSNMFCNNHAEAGCGAISAADAFSLQHNLYLFTAFENSSKNGCGAIQ-AQFSLSRNVSPI 192
Db 155 vsllsknftldngatltakltlletlglmsallseutskkggaltldlqngqv 214
QY 193 SFARNRADLNGAICCNLCGNNVNPFTGNSAT-----NGAICISDLNTS 242
Db 215 sfadnctsdsgaalftesavtlisnakvsfidihtvgassstlqdmgsaica---ykt 271
QY 243 EKGSLSLACNGEPTFASNAKEKGAIAKHMVLRNGPVSFINNS-----AKIGAIAT 297
Db 272 ldkvltlqgmllfsmcttaggalyvkkleasgqllfsmstnggtapkggalt 331
QY 298 QSGGSLSLIAGEGCVLPONNSORTSDGLVMAVLEKDALSSLEARNC-DLLFPPDV 356
Db 332 edsgelsadsqdivlglvtstlprlnrsldlgtaskmlaalsaaagraalfyfpil 391
QY 357 QESSS-----KESPLPSLSQASVSPTRATAPLVIOTSAKRSVIFSSERLSEEEKT 408
Db 392 tgsstvtvdkvnefpadsalqyt-----gnllfeyeklselea 432
QY 409 PD-NLTSOLQPIELKSGRLVLDRAVLSAPSLSDPQALLIMAGSLTSSDLKATL 467
Db 433 dsknltskllqprlsgltslkhyvltlqatfqadsstlemvgtltlepadstlnl 492
QY 468 SIPLHSLDTEKSVYIH---APNLISQKIFLSNSGDFEYENWELLSKQNNIPLTSL 522
Db 493 vlnissidgakkietkatskntlsgltllldplgtfeyenhlnrpqsdlllelka 552
QY 523 KEQSHLHPGNNLS--HFGYQGM--TFSMKDSDEGSLA--NWTFKNVYVHNERQSTL 577
Db 553 tvststvtcpdlmgckthyygqgtwpiw---gtgastatctfwkctgyipmperigs 609
QY 578 VANTLNTYSDMOAVQSMINTIAHG--GAYLFGTWGSAVSNLFYAHDSQGPIDNMHRS 635
Db 610 vpslwnafidissihymetaneqlqdrat--wcaglsnlt--hkdsklrrgtfnhs 665
QY 636 LGYLFGISLHSDHSEFLAAGQLLCKSSDSPTSTFTSYIAVVOQLATSLMKI---- 691
Db 666 ggyvlyggnlhtcskllisaafcglfgrdylfvakngqlyvggyllyqhmetylsplck 725
QY 692 -----SAQACYNESIHELKTYRSFSEKFGSMHSAVASEVCASLPIV 735
Db 726 rpslslyvpeipylfslgnlsythtdndlktkyltptvk-gsvngndsfalefgripic 784
QY 736 SNGGGLFSFSIFSLQGFQGTODGFEESSGEIRSFSSAFRNLSLPIGTFFEKSG-KT 794
Db 785 ldeaalteqymplmkidfvyahgeglkeqgleatefsgslvlnalrldlrdesdcqd 844
QY 795 RYYVYFLGAVYIDLKRDVEGPPVLLKNNAVSWAPMANLDSRAVMFRLLTQNALH-RLQT 853
Db 845 atynllyt-glytdlvrnptcdtllrlsgdskwltfgulardgalvllagphfctnfnfea 903
QY 854 LLNVSVLRGSHSYSLDLGTTTFE 878
Db 904 fsgfslrtgstrnynvdlgakygf 928

```

RESULT 14

Y69368
ID 169368 standard: Protein; 945 AA.

AC Y69368;

DT 19-JUN-2000 (first entry)

DE Amino acid sequence of the CPN100394 polypeptide.

KM CPN100394; Chlamydia infection; immune response; vaccine.

OS Chlamydia pneumoniae.

FH Key Location/Qualifiers

FT Peptide 1..43 /note="signal peptide"

PN W020001183-A2.

PD 02-MAR-2000.

PF 18-AUG-1999; 99MO-IB01449.

PR 20-AUG-1998; 98US-0097187.

PR 20-AUG-1998; 98US-0097188.

PR 20-AUG-1998; 98US-0097189.

PR 20-AUG-1998; 98US-0097190.

PR 20-AUG-1998; 98US-0097195.

PR 20-AUG-1998; 98US-0097196.

PR 20-AUG-1998; 98US-0097197.

PR 27-AUG-1998; 98US-0097191.

PR 17-AUG-1999; 99US-0376770.

PA (CONN-) CONNAUGHT LAB LTD.

PI Murdin AD, Oomen RP;

DR WPI; 2000-224703/19.

DR N-PSDB; Z61508.

XX Novel antigens and corresponding DNA molecules that can be used to

PT prevent, treat and diagnose disease caused by Chlamydia infection in

PS mammals, especially humans.

PS Claim 18; Fig 13A-F; 20pp; English.

XX Y69362-69 represent Chlamydia pneumoniae polypeptides. The polypeptides

CC are present in the bacterial membrane structure, in the external

CC vicinity of the membrane structure, in the inclusion membrane

CC structure, in the external vicinity of the infected cell. The polypeptides may be

CC used to prevent, treat and detect the presence of Chlamydia infection

CC and/or the presence of Chlamydia in a sample. The polypeptides may

CC also be used to induce an immune response in a mammal. The vaccine

CC vector comprising the polynucleotides is used to induce an immune

CC response in a mammal. Antibodies directed against the polypeptides

CC may also be used therapeutically to treat and/or prevent a Chlamydia

CC infection.

Sequence 945 AA:

Query Match 18.8%; Score 851; DB 21; Length 945;
Best Local Similarity 28.6%; Pred. No. 1.1e-62;
Matches 248; Conservative 153; Mismatches 367; Indels 98; Gaps 25;

```

74 SYCMFVSKLHITDPKELFKKGL-----STONFRFLSTFCSSKSSPSIIHQKNGQ 127
Db 117 sfdlislvitpskpsavttcgslvsgavqlqdlntlvltmsavdgglv--kynsc 174

```

ID Y94327 standard; Protein: 928 AA.
XX
AC Y94327;
XX
DT 11-AUG-2000 (first entry)
XX
DE Chlamydia pneumoniae 98kD putative outer membrane protein.
XX
KM Chlamydia; antigen; vaccine; infection; outer membrane protein.
XX
OS Chlamydia pneumoniae.
XX
PN MO200026237-A2.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99MO-CB03579.
XX
PR 29-OCT-1998; 98US-0106070.
XX
PR 01-MAR-1999; 99US-0122066.
XX
PR 27-OCT-1999; 99US-0428122.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Mordin AD, Oomen RP, Dunn PL;
XX
DR MPI: 2000-365569/31.
XX
DR N-PSDB: A27021.
XX
PT Novel Chlamydia 98 kDa putative outer membrane protein antigen, used
XX for vaccination and protection against Chlamydia infection .
XX
PS Claim 6; Fig 1; 93pp; English.
XX
CC The present sequence is the 98kDa putative outer membrane protein from
CC Chlamydia pneumoniae. The genomic sequence was amplified using two
CC PCR primers. The 5' primer contains a NotI restriction site, a ribosome
CC binding site, an initiation codon and a sequence close to the 5' end of
CC the 98kDa putative outer membrane protein coding sequence. The 3' primer
CC contains the sequence encoding the C-terminal sequence of the putative
CC outer membrane protein and a BspGI restriction site. The stop codon was
CC excluded and an additional nucleotide was inserted to obtain an in-frame
CC C-terminal fusion with the histidine tag. The PCR product was cloned
CC into a eukaryotic expression vector (pCA-Myc-His) by restricting both
CC the vector and the PCR product with NotI and BamHI and performing a
CC ligation reaction. This expression vector was injected intramuscularly
CC and intranasally into mice, which were subsequently inoculated with
CC Chlamydia pneumoniae. The chlamydial lung titers of the immunised mice
CC were lower than those of the controls. Thus the 98kDa putative outer
CC membrane protein can be used as a vaccine to provide protection against
CC Chlamydia infections, especially Chlamydia pneumoniae infections.
CC The present polypeptide may also be administered orally to treat
CC Chlamydia infection.
XX
SQ Sequence 928 AA:

Query Match 19.1%; Score 866; DB 21; Length 928;
Best Local Similarity 30.4%; Pred. No. 5.6e-64;
Matches 263; Conservative 133; Mismatches 365; Indels 114; Gaps 24;

QY 74 SYCMFVSKLHITDPKALFEKCDLSIONFRPLSTFDCSSKESSPSIIHOKNQLSRNN 133
DB 118 LFIgfssltfiaspgsiltlgkavs-----cst-----gslsltkn 154
QY 134 GMSMFCFNHAECSGSAISAFSLQHNHYLFTAFENSSKNGCALQ-AQTFSSLSRWSP 192
DB 155 vsllfskntscdggatctactlscgtmsaltsemtsskkgalqtsdaltltgqgev 214
QY 193 SFARNRADLNGAICCSNLICSGVNPDLFTGNSAT-----NGAICICISDLNTS 242
DB 215 sfadntssdsagaiftcaasvltisnmekvstfidhkvvgasssttgdmssgysaice---ykt 271

QY 243 EKGSLSLACNQTETLFASNSAKEKGAIYAKHMYLRYNGVFSFINNS-----AKIGCALAI 297
DB 272 tdtkvltcngmllfennsttagaivykkkleaaggltilfsmnsvngtapkgaaai 331
QY 298 QSGCSLIILAGESSVLFQNNNSQRTSDOGLVRNAYILEKRAILSSLEARNG-DILFFDP 356
DB 332 edsgeislssadsqdvlfntvtsttptgntsrldgtsakmcalrsaagralyfydpt 391
QY 357 QESSS-----KESPLPSSLOASVTSPTPATASPLVITQTSANRSVIFSRSLSEEXT 408
DB 392 tgsstvtvdkvnecpadaalqyt-----gnlltgeklseteaa 432
QY 409 PD-NLTSQIQPIELKSGRLVLDRAVLAPSLSODPOLLMFACTSLKTSDDLKATL 467
DB 433 dsknltskllqpvlsagtlslkhyvltlqgafitqgadsrlendvgttleppdstlml 492
QY 468 SIPHLSDTEKSTTH-----APNLSTQKIFLSNGSENYENVELSKQNNIPLTLS 522
DB 493 vniissidgakkakietatsknltlsgtcltlpctfyenhsirnpssydllelkaag 552
QY 523 KEQSHLPLPDGNLSS--HFQYQGDW-TFSMKDSDEGSLIA--NMTPKNYVPHPERQSTL 577
DB 553 tvstlavtpdplmgekfhygygctwplw---gtgaatcatcnvcktgylpmpetigs 609
QY 578 VANTLWNTYSDMQAVOSMINTIAHG--GAYLFGTWGSAVSNLFYAHDSGKPIDMNHRS 635
DB 610 vpsnlwnafidleslhylmelaneglgdraf--waglsnff--hkdstkrrgrfhis 665
QY 636 LGTLFGISTHSLDDHSPCLAAQGLLCKSSDFITSTETSYIATVQAQATSLMKI---- 691
DB 666 ggyvlygnlhtcsdkllsaefcqlfgdrdyvakqgvtvyggtlyghneyaislpckl 725
QY 692 -----SAQACYNESIHELKTKYRSRSGKGFMSHVAASGEVCASIPIV 735
DB 726 rpsclsyvptelpvlisgnlsyhtndlktlytlypvk-gswgndstalefgtgrapic 784
QY 736 SNGSGLFSSFSIFSKIQGFSGTODGFEBSGGELIRFSASSFRNISLPIGITEKKSQ-KT 794
DB 785 ldesalfegymplmkqifvyahqegfkeqgtearefsgsrllvalalpigrfdkesdcq 844
QY 795 RTYYRFLGAYIODLKRDESVGPVLLKNVSWDAPRANLDSRAYMRRLNQRALH-RLOT 853
DB 845 atynlclg-yvvdilvrnsdpdcttlrlfsgdswktlgtlnlargalvltagnhcfnsfnea 903
QY 854 LNWSCVLNRQSHSYSLDGLTTRYF 878
DB 904 fsqfslfegsrnyndlgakqyf 928
RFSULT 12
ID Y69369
AC Y69369 standard; Protein: 918 AA.
XX
XX Y69369;
XX
DT 19-JUN-2000 (first entry)
XX
DE Amino acid sequence of the CPN100395 polypeptide.
XX
KW CPN100395; Chlamydia infection; immune response; vaccine.
XX
OS Chlamydia pneumoniae.
XX
PN MO200011183-A2.
XX
PD 02-MAR-2000.
XX
PF 18-AUG-1999; 99MO-IB01449.
XX
PR 20-AUG-1998; 98US-0097187.
XX
PR 20-AUG-1998; 98US-0097188.
XX
PR 20-AUG-1998; 98US-0097189.
XX
PR 20-AUG-1998; 98US-0097190.

```

Db      349 nltltsapcstrtnaiylgssakitniraagqsiylfydplaasntgagavl----- 400
Qy      375 TSPPTATASPLVIQTSTANRSVIFSSERLS--EEEKTPDNLTSQLOQPIELKSGRLVLDRA 433
Db      401 tlnqpdnsnpl-----dysgvlvfgsklsadeakaadnftslkqplalasgtlalkgnv 456
Qy      434 VLSAPSLSDPPQALLIMEAGSLKTSDD-LKLAATLSLPLSLDMEKSVTHAP-----N 486
Db      457 eldvngftqlegstllmqpgtklkadeaislklvldlalegknsvslategankltc 516
Qy      487 LSIQKIFLNSGDENFYENVELLSKEQNNIPLLTLSKEQSHLHPDGNLSS-----HF 539
Db      517 ltsplvlfqdsq--nfyeshtlmgafcpqlvlfvaataasdiyl-dalltsqpqtpephy 573
Qy      540 GYCGDMTFSSKSDDEGSLIANWPKNYVPHPEROSTLVANTLWNTYSDMQAVOSMINTI 599
Db      574 gygghweatwadetsaksgtmctwtgynperrasvvpdsfwaastfdlrltqlqintsq 633
Qy      600 AHGCAVLFCTWGSVSNLFLFAHDSGKPRIDNMHRSGLYFGISTHLDHSFCLAGOL 659
Db      634 ansllyqgrglwaagtanfth-kdksqtn-qaftrkxygylvggsaaedfseailfvalcql 691
Qy      660 LCKSSDSFTSTETTSYIATVQAO-----LATSLMK-----ISAQACYNES 700
Db      692 fykdldlfienvcsnhyjaelyqhrafqglmpstfgstldmkdipil lnaqjlsysyc 751
Qy      701 IHEKITYRSFSEKGFSGSHSVAVSGEVCASIPV-VSNGSGLSSFSFESKLGQFGSGTOD 759
Db      752 kmnditrysy-peaagswtensgalelqgsalylypkcapffqgyfplkfqavysrty 810
Qy      760 GFPESSSEISFSASSFRNLSLPIGTFEKKSKOKTRFYFFYFIAGYIODLKRDVESCHVVL 819
Db      811 nlfksgaeatadgdglvncslpqrllleklsedekmfelslayldqvyfknprfetsl 870
Qy      820 LKNAVSDAPMANLDSRAYMFLTNORALH-RLQTLNVSCVLKQSHSYSLDICTTYRF 878
Db      871 mvsgaswtscldknlarqafilasagshltlsphealsgeaayelrghshlynvdcgllysf 930

RESULT
W88424
ID      W88424 standard; Protein; 930 AA.
XX
AC      W88424:
XX
DT      26-APR-1999 (first entry).
XX
DE      Chlamydia pneumoniae surface exposed protein Omp11.
XX
KM      Omp11: outer membrane protein 11: surface exposed protein:
XX      antigen; infection; diagnosis; vaccine; atherosclerosis; asthma.
XX
OS      Chlamydia pneumoniae.
XX
PN      W09858953-A2.
XX
PD      30-DEC-1998.
XX
PF      19-JUN-1998: 98WO-DK00266.
XX
PR      23-JUN-1997: 97DK-0000744.
XX
PA      (BIRK.) BIRKELUND S.
XX      (CHRI.) CHRISTIANSEN G.
XX
PI      Birkelund S, Christiansen G, Knudsen K, Madsen A;
XX      mydind P;
XX
DR      WPI: 1999-105610/09.
XX      N-PSDB: X06823.
XX
PT      Species-specific test for identifying mammals infected with

```

```

PT      Chlamydia pneumoniae - comprises detecting antibodies specific for
PT      outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT      these proteins
XX
PS      Claim 7: Page 63-65; 115pp; English.
XX
CC      This polypeptide comprises the novel 97.6 kDa surface exposed
CC      protein Omp11 of the human respiratory pathogen Chlamydia
CC      pneumoniae. Its amino acid sequence was deduced from DNA (see
CC      X06823). Isolated from a C. pneumoniae expression library. The
CC      invention provides 12 novel surface exposed proteins, Omp4-Omp13
CC      (see W88417-28), and nucleic acid sequences encoding them (see
CC      X06816-27). A new species specific test is claimed that is used
CC      to identify mammals (including humans) infected with Chlamydia
CC      pneumoniae. The test comprises detecting antibodies specific for
CC      Omp4-Omp13 or detecting nucleic acid fragments encoding these outer
CC      membrane proteins, especially by PCR. The proteins are also used
CC      in the diagnosis of C. pneumoniae infection in mammals. The
CC      nucleic acids and proteins can also be used in the immunization of
CC      mammals, the nucleic acids being particularly useful as DNA
CC      vaccines for effecting in vivo expression of antigens. The
CC      vaccines may also prevent atherosclerosis and bronchial asthma,
CC      which are possibly associated with C. pneumoniae.
XX
SO      Sequence 930 AA:

Query Match      20.0%; Score 906.5; DB 20; Length 930;
Best Local Similarity 30.7%; Pred. No. 2,1e-67;
Matches 258; Conservative 150; Mismatches 345; Indels 87; Gaps 26;

Qy      98 LSIQNFPLSF-----TDCSSKSSPSIIHQKNGOLSRNNGSMSPCR--NHAEGSCGA 149
Db      119 ltfefgnlffiaapgtlvagysktsl-----segalltdngllflsqvnsaannqga 174
Qy      150 ISADAFSLQHNLYLFTAEENSCKNGCAI-QAOTFSLSRNVSPISFARNRADLNGAICG 208
Db      175 ltkctlsisngtsstlftnsakklgyaelyssaaslsqgtglvfmmnkgcggagall-- 232
Qy      209 SNLICSGNV---NPLFTGNSATN---GGAICISDLNTSEKGSLSLACQETLRASNS 261
Db      233 -gfeassstlcqnsllffsgntataaqqgalyc--ekgetpdltlisgnksaltfaens 288
Qy      262 AKEKCAIVAKHMYLRYNCGVPIFN-----SAKIGCALIOSGSGSLIAGGSYLFON 316
Db      289 svcggaicahglidsaaprlfsmrcgntaaqgagaladagslsiaangdflfig 348
Qy      317 NS-ORTSDQGLVRNAIYLEKDAISSLARNG-DLFFDPVVOESSSEKSLPSSLOASV 374
Db      349 nltltsapcstrtnaiylgssakitniraagqsiylfydplaasntgagavl----- 400
Qy      375 TSPPTATASPLVIQTSTANRSVIFSSERLS--EEEKTPDNLTSQLOQPIELKSGRLVLDRA 433
Db      401 tlnqpdnsnpl-----dysgvlvfgsklsadeakaadnftslkqplalasgtlalkgnv 456
Qy      434 VLSAPSLSDPPQALLIMEAGSLKTSDD-LKLAATLSLPLSLDMEKSVTHAP-----N 486
Db      457 eldvngftqlegstllmqpgtklkadeaislklvldlalegknsvslategankltc 516
Qy      487 LSIQKIFLNSGDENFYENVELLSKEQNNIPLLTLSKEQSHLHPDGNLSS-----HF 539
Db      517 ltsplvlfqdsq--nfyeshtlmgafcpqlvlfvaataasdiyl-dalltsqpqtpephy 573
Qy      540 GYCGDMTFSSKSDDEGSLIANWPKNYVPHPEROSTLVANTLWNTYSDMQAVOSMINTI 599
Db      574 gygghweatwadetsaksgtmctwtgynperrasvvpdsfwaastfdlrltqlqintsq 633
Qy      600 AHGCAVLFCTWGSVSNLFLFAHDSGKPRIDNMHRSGLYFGISTHLDHSFCLAGOL 659
Db      634 ansllyqgrglwaagtanfth-kdksqtn-qaftrkxygylvggsaaedfseailfvalcql 691
Qy      660 LCKSSDSFTSTETTSYIATVQAO-----LATSLMK-----ISAQACYNES 700

```

[illegible]

D6
628 valkvrsqgctrlgwceglsnltf--hkdsuklnkgrfnhsagyyvgatalltaadnlltaa 665

OY
656 AGQLLCKSSDSOFFTSTETTSYIATVQAQ-LA?-----SLMK-----ISAQACYN 698

D6
686 fccqlfygdrcdhflinkrtasayaaailhlphlaetlspslrlylpgeesepvlfdaqlsytl 745

OY
699 ESHIELTKTRRPSKKEGFGSMHVAWNGVCVASIFIVS-NMSGLEFSSSRISFKIQGRSGT 757

D6
746 ysktmtttcylvtgapk-gesswyndgcacalelaasilphcalnsheglthayfprikveasylh 804

OY
758 QDGEE-SSGEIRSFSASSFERNISLPDITGITFEKKSQKTRTYVFLGAYIDPLKRDVESEGP 816

D6
805 qdsikercntlvrsfdsgdlinwsvpiigltferfarnerasyeatvlvyadvyrknpcdt 864

OY
817 VLLKNVMSDMADMANLDNSA-----YMERLTNQALRLRIQTLLNVCVLRGOSHRY 868

D6
865 tallimtlswkttgtnlsrqaglgtagltyalts-----pnlevtsnlsmeltrgsrsry 917

OY
869 SLDLGTTRYR 878

D6
918 nadlgkgkf 927

RESULT 7

Y35052

I1D Y35052 standard; Protein: 930 AA.

XX AC Y35052;

XX DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae surface exposed polypeptide.

XX XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis

KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

XX XX Chlamydia pneumoniae.

OS W09927105-A2.

PND 03-JUN-1999.

PD 20-NOV-1998; 98WO-1B01890.

XX PF 04-NOV-1998; 98US-O107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX XX (GENSET) GENSET.

PA Griffiths R;

XX PI MPI, 1999-357842/30.

DR Genome sequence of Chlamydia pneumoniae

FT Page 940-942; Disclosure: 1912pp; English.

XX PS Y3584-Y35879 represent the proteins encoded by all the open reading

CC frames in the complete genome (see X01990) of Chlamydia pneumoniae.

CC C. pneumoniae causes respiratory disease such as pneumonia and

CC bronchitis and is thought to be a contributing factor in heart

CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema

CC nodosum or pharyngitis. The polypeptides encoded by the open reading

CC frames of the C. pneumoniae genome (see Y3584-Y35879) can be used in

CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae

CC nucleotide sequences can also be used as immunogenic compositions,

CC especially where the vector directs the expression of a neutralising

CC epitope of C. pneumoniae.

XX XX Sequence 930 AA:

XO

YY	367	PSLSQASTPTSPPTATPLVIOFSNARSVLFSSERLSEEEKPDDNLTGSLQOLPTELEKSGR	426
Db	339	PSLSQASTPTSPPTATPLVIOFSNARSVLFSSERLSEEEKPDDNLTGSLQOLPTELEKSGR	398
QY	427	LVLDRAVLAPSLSODPQALLIMEAGTSLKTSDDLKL	464
Db	399	LVLDRAVLAPSLSODPQALLIMEAGTSLKTSDDLKL	436
RESULT	4		
Y90239			
ID	Y90239	standard; Protein: 928 AA.	
AC	Y90239;		
XX			
DT	29-AUG-2000	(first entry)	
XX			
DE	Chlamydia antigen CPN100638.		
XX			
KW	Chlamydia antigen; diagnosis; infection; community acquired pneumonia;		
RW	therapy; upper respiratory tract disease; bronchitis; sinusitis;		
XX	asthmatic bronchitis; adult-onset asthma; acute exacerbations of asthma.		
XX			
OS	Chlamydia pneumoniae.		
XX			
PN	MO200032794-A2.		
PD	08-JUN-2000.		
XX			
PE	01-DEC-1999;	99MO-CA01147.	
XX			
PR	01-DEC-1998;	98US-0110339.	
PR	01-DEC-1998;	98US-0110340.	
PR	01-DEC-1998;	98US-0110427.	
PR	01-DEC-1998;	98US-0110428.	
PR	01-DEC-1998;	98US-0110438.	
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
XX			
PI	Murdin AD, Oomen RP, Wang J;		
XX			
DR	WPI: 2000-412339/35.		
DR	N-PSDB: A30851, A30852.		
XX			
PT	Nucleic acids encoding polypeptide antigens from Chlamydia useful for		
PT	preventing, diagnosing and treating diseases such as community acquired		
PT	pneumonia, bronchitis, sinusitis and asthmatic bronchitis, adult-onset		
PT	asthma.		
XX			
PS	Claim 16; Fig 5; 174pp; English.		
XX			
CC	This sequence is a Chlamydia antigen of the invention, designated		
CC	CPN100638. The nucleic acids (and their complementary sequences) may be		
CC	used as diagnostic agents for detecting the presence of nucleic acids		
CC	encoding Chlamydia antigens in samples according to standard methods,		
CC	and therefore, for diagnosing Chlamydia infections. For example, they may		
CC	be used as primers and probes for diagnostic polymerase chain reaction		
CC	(PCR) assays. Antisense sequences may be used to down regulate		
CC	expression of the proteins and may be used to treat infections. The		
CC	nucleic acids may also be used to produce the protein antigens they		
CC	encode according to standard recombinant DNA methodologies. The		
CC	proteins may then be used as antigens for the production of antibodies		
CC	(i.e. as vaccines) for preventing infection by Chlamydia. The		
CC	antibodies may also be used as diagnostic reagents for detecting		
CC	infections. Chlamydia is a pathogen implicated in the development of		
CC	(for example) community acquired pneumonia, upper respiratory tract		
CC	disease (especially bronchitis and sinusitis, asthmatic bronchitis,		
CC	adult-onset asthma and acute exacerbations of asthma in adults.		
XX			
Sequence	928 AA:		

Query Match 20.8%; Score 944.5; DB 21; Length 928;

[illegible]

AC W88423;

PI Probst P, Bhatia A, Skeiky YAM, Fling SP, Jen S, Stromberg EJ:
 XX
 XX WPI: 2000-431303/37.

PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 XX comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence
 XX
 PS Claim 2: Pages 179-181; 256pp; English.

CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumonia is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the
 CC present invention.

XX Sequence 880 AA:

Query Match 99.6%; Score 4513; DB 21; Length 880;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 874; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MRPDHNFFCCCLCAALISSRAVLFGDPLGEPVALTKPNPHVOTFFEDCMESLPALCA 60
 DB |||||||
 QY 3 MPDMMFFCCLCAALISSRAVLFGDPLGEPVALTKPNPHVOTFFEDCMESLPALCA 62
 DB |||||||
 QY 61 HASODPLVVLGNSYCFMFSKLTHTDPKALFEKSGDLSIQNFPLFTDCSSKESPSI 120
 DB |||||||
 DB 63 HASQDPLVVLGNSYCFMFSKLTHTDPKALFEKSGDLSIQNFPLFTDCSSKESPSI 122
 DB |||||||
 QY 121 IHOKGOLSLRNNGSMFCCNHAEGSGAISADAFSLQHNLYLFTAFEBNNSKNGCAIOA 180
 DB |||||||
 DB 123 IHQKNQGLSLRNNGSMFCCNHAEGSGAISADAFSLQHNLYLFTAFEBNNSKNGCAIOA 182
 DB |||||||
 QY 181 QPFSLSRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFPTGNSATNGAICCSIDLN 240
 DB |||||||
 DB 183 QTSLSIRNVSPISFARNRADLNGAICCSNLICSGVNPPLFFPTGNSATNGAICCSIDLN 242
 DB |||||||
 QY 241 TSEKGSLSLACNDETLFASNSAKEKGAIYAKHMLRYNPGVPSFINNSAKIGGAIOAG 300
 DB |||||||
 DB 243 TSEKGSLSLACNDETLFASNSAKEKGAIYAKHMLRYNPGVPSFINNSAKIGGAIOAG 302
 DB |||||||
 QY 301 GSLSLIAGEGSVLPQNNQRTSDQGLVRNAIYLEKDAIISLSEARNGDILFDPIVOESS 360
 DB |||||||
 DB 303 GSLSIAGEGSVLPQNNQRTSDQGLVRNAIYLEKDAIISLSEARNGDILFDPIVOESS 362
 DB |||||||
 QY 361 SKSPSPSSLOAVSTPTPTASPLVLOTANSRVIFSSRSLSEETPTNLSLOQOP 420
 DB |||||||
 DB 363 SKSPSPSSLOAVSTPTPTASPLVLOTANSRVIFSSRSLSEETPTNLSLOQOP 422
 DB |||||||
 QY 421 ELKSGRLVLDKRAVLNAPSLSDPQALLMEAGTSLKTSDDLKATLSIFLMSLDTERKV 480
 DB |||||||
 DB 423 ELKSGRLVLDKRAVLNAPSLSDPQALLMEAGTSLKTSDDLKATLSIFLMSLDTERKV 482
 DB |||||||
 QY 481 TTAAPNLISOKITLNSGNDNFENVELSKEDONIPFLTLSEQSHLHPDGNLSHF 540
 DB |||||||
 DB 483 TTAAPNLISOKITLNSGNDNFENVELSKEDONIPFLTLSEQSHLHPDGNLSHF 542
 DB |||||||
 QY 541 YGDDWTFMSKDSDEGSLANWTPKNVPHPEPOSTLVANTLWNTYSDMOAVOSMINTIA 600
 DB |||||||
 DB 543 YGDDWTFMSKDSDEGSLANWTPKNVPHPEPOSTLVANTLWNTYSDMOAVOSMINTIA 602
 DB |||||||
 QY 601 HGGAYLFGTGSVAVSNLFVAHDSGKPIDMWHHRSICGLFGISTHSLDDHFCILAGQL 660
 DB |||||||

DB 603 HGGAYLFGTGSVAVSNLFVHDSGKPIDMWHHRSICGLFGISTHSLDDHFCILAGQL 662
 DB |||||||
 QY 661 GKSDDSFITSTETSYIAIVAOALATSLMKISAQACYNNSHELKTKVRSKRGSGWH 720
 DB |||||||
 DB 663 GKSDSFITSTETSYIAIVAOALATSLMKISAQACYNNSHELKTKVRSKRGSGWH 722
 DB |||||||
 QY 721 SVAISGEVCASTPIYVNSGGLFSFISFISKLOGFSGTOGFEESSGEIRSFASSFRMS 780
 DB |||||||
 DB 723 SVAISGEVCASTPIYVNSGGLFSFISFISKLOGFSGTOGFEESSGEIRSFASSFRMS 782
 DB |||||||
 QY 781 LPGITFEKRSQKTRTYVFLGAYTDLKRDYESGPPVLLKNVSWDAPANLDSRAYWF 840
 DB |||||||
 DB 783 LPGITFEKRSQKTRTYVFLGAYTDLKRDYESGPPVLLKNVSWDAPANLDSRAYWF 842
 DB |||||||
 QY 841 RLTONRALRLQTLNLVSCVLGQSHSYSLDGTYYRF 878
 DB |||||||
 DB 843 RLTONRALRLQTLNLVSCVLGQSHSYSLDGTYYRF 880
 DB |||||||

RESULT 2
 ID B13638 standard; Protein: 866 AA.
 B13638
 AC B13638:
 XX
 XX
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE C. trachomatis pmpl gene amino terminus minus signal sequence protein.
 XX
 KW Chlamydial infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 OS Chlamydia trachomatis.
 OS
 PN WO200034483-A2.
 PD 15-JUN-2000.
 XX
 XX 08-DEC-1999; 99WO-US29012.
 PF
 XX 08-DEC-1998; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 PA
 XX
 XX
 DB WPI: 2000-431303/37.
 PI Probst P, Bhatia A, Skeiky YAM, Fling SP, Jen S, Stromberg EJ:
 XX
 XX WPI: 2000-431303/37.

PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence
 PS Claim 2: Pages 205-207; 256pp; English.

CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumonia is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689.6
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000
TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-23

Query Match 2.88; Score 125; DB 2; Length 1566;
Best Local Similarity 19.58; Pred. No. 0.048;
Matches 172; Conservative 109; Mismatches 315; Indels 288; Gaps 40;

QY 11 LCAAILSTPAVL---EGDPLGETALLTKPN---NHVCTFFEDCTMESLFPALCAHASQ 64
DB 18 LCGALL-GTALLSVGOKLAETSTSTSGVNTAVGT-----ETGNPATNLPDKQ 69
QY 65 DDEPLYLGNSYCFVSKLHTDPKREALFEKGLDSTIONFRFLSTDCSSKSSPSIITHQK 124
DB 70 DNP-----SSQAEFSQAGQAK 86
QY 125 NGQLSLR-NGSMSCFNHAGSGAISADA-----FSLQHNLYLFTAFEESSSKNG 175
DB 87 TGMASVDVSTSEIDBAKSAQENGVTVSODATYDKGTETSDENOKETELKDDYSK--- 143
QY 176 GAIOAQTFSLSRNVSPISFARNRADLNGAICCSNLICSGNVNPLFTGNATNGAICC 235
DB 144 ---QADDIQKTEDYKAAVARNOAETD-----RITQENAKKAQYEQDLANKKAVER 193
QY 236 ISDLNTSEKG--SLSLACNOETLFASNSAKEKGAIY--AKHNVLRNGPVSTINNSAK- 290
DB 194 ITNENQQRKADYBAKLAQYOKDLAAVOQANNDSQAAYAAKAEADKELARVOAANAATAKK 253
QY 291 -IGAITAIOGSGSILAGSGSVLFQNNQRTSPQGLVRNAIYLEKDAIILSSLEARNDI 349
DB 254 EYEBALANTTKNEQIKA-ENVAIQQRNAQAKADYE-AKLAQY-EKDL---AAQSGNA 306
QY 350 LFPDPIVOESSKSPPLSSIOASVTSPTPATASPLVIGTSANRSVIFSSERLSEER-KT 408
DB 307 TNENDYQAKKAAYQEL-ARYQAANAQAQAYEQALANTAKNAQITAENEAIIQQRNAQA 365
QY 409 PDNLTSQLQ---PIELKSGRLV-----LKDRAVLSAPSL----QD 443
DB 366 KANEYAKLAQYOKDLAAOQANANEAADYQEKLAAYEKELARVOAANAQAQYEQKVOE 425
QY 444 PQA-----LLIMEAGTSLKTSDDL------AT 466
DB 426 ANAKNAETTEANRAIRERNAKAKTDYELKLSKYOELAQYKKDLAEYPAKLAQYODEQA 485
QY 467 LSLPLHSLDTEKS-----VTIHPNLSTQ-----KIFLSNGDENYENVEL 508
DB 486 IKALEELEKHKHEDMWLSEPSAQSLVYDLBNQOISLVTDWKLLKASLDESFSHDTEQ 545
QY 509 LSK---EONNIPPLTLSEKQSHLHPDGNLSSHFGYQGDWTFESMKDDE-----G 555
DB 546 YNKHNLQPDNINITYL--EQADVASSVEIFGNPGDKRAGMTTIVSNGSEYKFAVLLKRG 603

QY 556 HSLIANMTPKKNYPHBEROSTLVANTLWNTYSDMQAVQSMINTIA-----HCGAY 605
DB 604 QSNATATYT-----NLKNSYNGKKISKVVYKTYDPDSKQNPPTGNW 646
QY 606 L-----FGTWGA-----VSNLFYAHDSCKPID-----NMHRS 635
DB 647 LGIFDPTLGVEFASAVTGONKDTISIFIKNEFTFYDEDGNPIDFDNALSVASLNEHNS 706
QY 636 LGTLFISTHSDHS--FCLAAQGLGKSSDSFISTETTSYIATVOAQLATSLMKISA 693
DB 707 I-----EMAKDYSGTGVKISGSSIGE-KNGMIYATDTLNF----- 740
QY 694 QACYNESIHLEKTKYRSFSKEGFG-----SMH---SVAVSG 726
DB 741 KKEGGSLH---TMYTRASEPGSGHDSADAPNSWYGAGAVRMSG 781

Search completed: May 6, 2001, 19:25:45
Job time: 280 sec

Query Match 2.8%; Score 129; DB 2; Length 1600;
 Best Local Similarity 18.3%; Pred. No. 0.021;
 Matches 149; Conservative 107; Mismatches 275; Indels 284; Gaps 36;

73 NSYCFVSLHTDPKELFKKSG-----DLSTIONFRLSTDCSSKESPSIIHOKNG 126
 143 NOCVFLINMGITIGKDALINMGFTASTLIDISNENIKARNFTLQYDKALAEI-VNHG 201
 127 QLSLNNNSMSPFCRRHHAEGCGAISADAFSLQHNLYLFAFEENSSKNGCAIQ---AQTF 183
 202 LITVCKDGSVNI-----GGKVK-----NEGVIISVNGSISLLAGQKI 239
 184 SLGRVSP-----ISFARRADLNGAICCSNLICSGANNPLFPNGNSATNGAICCID 238
 240 TTSIDILNPTITYSIAAPEMEA-INLQDIFAK-----CGNIN-----VRA 277
 239 LNTSEKGSILACNOETLFASSNAKEKGAIV-AKHMYLRYNGPYFINSNAK----- 290
 278 ATIRNKGKLS-----ADSVSKDKSGNIVLSAKEGLAEIGCVISAQNOAKGCKLMI 328
 291 ICGAIAIQSGCSLSILAGEGSVLPONNSQRTSDQGLVNAIYLEKDALISLEARN----- 346
 329 TQDKVTLKGAVIDLSGREGGYLGGDER-----GEGKNGIQAKKTKLEKSTINVSCK 384
 347 -----GDLFPPIVQESS-----361
 385 EKGRAIVMGDIALIDGINIAOGSDIAKTGCFVETSGHDLSCIDDVIVDAKEWLLDDDDV 444
 362 -----KESPLPSLQASVTSPTASPL-----VIQT 389
 445 SLETTLSGRNNGENOCYTTCOTKRSPPKN-----SISKPT-LTNSTLQDLKRGSVNI 499
 390 SANRSVIFSSERLSEEEKTPDNLTSOLOQPIELKSGRLV--KRAVLSAPSLSDPOL 447
 500 TANNRIYVNS-----INLSNGSLTLHTRKDDGVKINGDITSENGN 540
 448 LINEACTSLKSSDLKATLISIPHSIDT-----EKSTIHAAPLISIQKIFLSNGDE 500
 541 LTKAKSMDVHNKNTLGTGFLNIVAGDSVAFEREGDKANATDAQTACGTTVANKDDK 600
 501 NF-YEAVELLSEKQNNPLTLTSEKSHLILPDGNSLSHFGYOGDMTFS-----WK 550
 601 QEFNNVSLNGCTKG--LKFIANQNNFTHKFPGCEIN-----ISGIYIINTTKKDYKWN 653
 551 DSDGHSILANMTPKNVVPHPEROSTLVANTLM-----NTYSQMAVQSMJNTIAHG 602
 654 ASKDSY-----W-----NVSLLTNTYQKFTFIKFDVDSGSMGQDLRSSRSFA-- 696
 603 GAVLFGTWSAVSNL-----FYAH-----DSSCKPIDNMH 632
 697 CVHFNGICGKTNNFICANAKALFKLPNAATDPKKELPITFNANITATVCGNSDSSVMEFDIH 756
 633 HRSGLVFGTSTHLDHDSCLAGOLGKSSDSFITSTETTSYIATVQALATSLMKIS 692
 757 ANLTSRAAGINMDSIN-----ITGGL-----DESIISHKNSNAFIRKIDLTT-----IN 800
 693 AOCYNESIHELKTR-YRSPSKGEGGSMHSAVAVSG 726
 801 ATGS-NFSLKOTKDSFYNEYSKHAINSSHNLITLG 834

RESULT 23
 PCT-US95-10661A-5
 Sequence 5, Application PC/TUS9510661A
 GENERAL INFORMATION:
 APPLICANT: Washington University, et al.
 TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fleh, Hohnbach, Teal, Albritton & Herbert
 STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/10661A
 FILING DATE: 16-AUG-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/296,791
 FILING DATE: 25-AUG-1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Treccartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: FP-59941/RFT
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1702 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 PCT-US95-10661A-5

Query Match 2.8%; Score 127.5; DB 5; Length 1702;
 Best Local Similarity 20.1%; Pred. No. 0.033;
 Matches 182; Conservative 119; Mismatches 296; Indels 309; Gaps 50;

66 DPL-----YVLGNSYCFVSKLHTDPKELFKKGD-LSTIONFRLSTDCSS----- 113
 277 DPLTNVAVLQDS-----GSLPYVD-----REKGMVLPFGSYFMAVGNKMSQENNIYK 326
 114 -----KESPSIIHOKNGOLSLRNNNGSMFRCRHHAEKSGAISADAFSLQHNLYLT 164
 327 PEFKATVLDKDTAGSLI-GSNTQYNNMPTGKTSVISNGSE-----SLNVDLFD-----S 374
 165 AFENSSKNGCAIQNOTFELSNNVSPISFARRADLNG-GAICCSNLICSGVNPPLFFT 223
 375 SQDTDSKKNHG-----KSVTLRSGGTLTLNNNDQG-AGGLFEE 413
 414 GQYEVKG-----TSDSTTWKAGVSA-----DQKTVTKVHNPKS 449
 224 GNSATNGAICISDINTSEKGSILACNOETLFASSNAKEKGAIVAKHMYLRYNGPV 283
 284 FINNSAKTGAIAIQG-----GSLTLAGEGVLFQ-----NNSORTSDQ-GLV--RNA 330
 450 --DLRAKIGKGTILVEGKNGKSLKV--GDGVYILKQADANNKYVAFSQVQIVISGRST 505
 331 IYLEKDAIIL-----SLEARNGLILFPDPI-----VOESSKESPLPSSL 370
 506 VVLNDDKQVDPNSTYFGFGRGRDLA--NGNNLTFEHNINDDGARLVNHNSTSTVITG 564
 371 QASVTSPTATASPLVIQTSANRSVIFSSERLSEEEKTPDNLTSOLOQPIELSGRLVK 430
 565 ESLTIDENIT-----TPYN-----578
 431 DRAVLAPSLSODPOLLINEAGTSLKTSDDLKATLISIPHSIDTEKSVTHAPNLSIQ 490
 579 ----IDAPD-EDNRYAF-----RRIKDGGLYLNLENTYVALRKGASTRSELPL----- 622
 491 KIFLSNGSDENFENVNELLSEKQNNPLTLTSLKQSHLILPDGVLSSHFGYOGDMTFSWK 550
 623 ----KNSGESH--ENMLYMKTSD-----AAKKNVNMHNINEMNNGFNIGTGE----- 664

QY 462 LKATLSIPLHS---LDTEKSV---THAPNLSIOKIFLSN-----SGDENFENVELLSKE 512
Db 1195 IFLTTLS-NVHENGRIHQEKNIOEIEKEALIEKVYVLPVHATGSKNFLKDLILGTR 1253
QY 513 QN-----NIPLLTLKESQSHLHPDGNLSHFGYQDMTFSMKDSDEGHSLIA 560
Db 1254 QNLSLVEVHPVLQNTSINNSTVQIHM-----EHFRR-----KDETNSEGLV 1301
QY 561 NMTPKNVHPERQSTLVANTLMNTYSDMQAVOSMINTIAHGAVLFGTWGSAVSNLFYA 620
Db 1302 NKTR-----MVKN-----YPSQKNITTORSKRALGOFRLSTQW---LKTINCS 1342
QY 621 HDSSGKPIDNMHHSLGCLFGLSTHSLDHSFCLAGQLGKSSDSFTSTETTSYATV 680
Db 1343 TQCIKQID--HSEKMKF--ITKSSLSDSSVTKSTQW---NSDSHIVTSAFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHELKTR 707
Db 1397 RSPFQNKFSHVQA-----SSYIDFKTR 1419

RESULT 20
PCT-US94-13200-6
Sequence 6, Application PC/TUS9413200
GENERAL INFORMATION:
APPLICANT: Emory University
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/13200
FILING DATE: 15-NOV-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Padst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU106CIP(2)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lach, D.
AUTHORS: Galschier, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993

RELEVANT RESIDUES IN SEQ ID NO: 6, FROM 1 TO 2319
PCT-US94-13200-6

Query Match 2.9%; Score 132.5; DB 5; Length 2319;
Best Local Similarity 20.2%; Pred. No. 0.02;
Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FECT-----MESIFPAL-----CAHASQDDPLVIGNSCWEVSKLHTDPREALF 92
Db 775 FKDSITPKNOMEKIEPOFEIEALMKVQSVSVSMMLLQCSHP-TPHGFILSDQGEAIV 833
QY 93 K-----EKDLSIQNFRFLFTDCSSKESPSTIIQKQSLRNNGSASF 139
Db 834 EALHDSHPAIDNSBSPSVTQLR-----PSHHSEKIVTTPQGLQLRKNSL--- 883
QY 140 RNHAEGSGAISADAFSLQHNLYFT-----AEE-----NSSKNGCAI 178
Db 884 ETTEVYKMKLGLQVSLPNSLMTTILSDNLKATFEKTDSSGFPDMPIVASSKLSSTAF 943
QY 179 QAOTFSLRNVSPISEFARNADLNGCAICCSNLICSGVNPFLFTGNSATNGAICCI 238
Db 944 CKRAYSLVCSHVPNLNASENSD-----SNILDS---TLMYSQSLPRDNLSTEND 991
QY 239 LNTSEK--GSLSLACNQTLEFASNAKEKGATY---AKHMYLRNCPVSFINNSAKIG 293
Db 992 RLREKRFHCIALTYDNTLEKQVSLMKTKNTKNTSTTEKLTSPSTIENSTQLOD 1051
QY 294 A1-----AIOGCSLILAGEGYLFEQNSQRTSDQGLVFNATYLEKDALIS--SLEARNG 347
Db 1052 AILKVNSEIQ---EVTLLINDGTLGKNST-----YLRNHLNLTSTSTK 1095
QY 348 DILF---FDPYVQSSSKESPLVSLQASVTSPTPATASPLVLIQTSANRSVIFSSERLSE 404
Db 1096 DIFHRKDEDPIPODEENTIMPFSKMLFLS-----ESSNMFKKTJNGNSL---NSE 1142
QY 405 EERTPDNLTSOLOPIELKSGRLVKRAVLASAP---SLSDQALLIMAGTSKLTSSD 461
Db 1143 QHSPKQVLYLM-----FKKYVKNQSLSEKKNKYVQDGFNTKIGLKDAFPNNMS 1194
QY 462 LKATLSIPLHS---LDTEKSV---THAPNLSIOKIFLSN-----SGDENFENVELLSKE 512
Db 1195 IFLTTLS-NVHENGRIHQEKNIOEIEKEALIEKVYVLPVHATGSKNFLKDLILGTR 1253
QY 513 QN-----NIPLLTLKESQSHLHPDGNLSHFGYQDMTFSMKDSDEGHSLIA 560
Db 1254 QNLSLVEVHPVLQNTSINNSTVQIHM-----EHFRR-----KDETNSEGLV 1301
QY 561 NMTPKNVHPERQSTLVANTLMNTYSDMQAVOSMINTIAHGAVLFGTWGSAVSNLFYA 620
Db 1302 NKTR-----MVKN-----YPSQKNITTORSKRALGOFRLSTQW---LKTINCS 1342
QY 621 HDSSGKPIDNMHHSLGCLFGLSTHSLDHSFCLAGQLGKSSDSFTSTETTSYATV 680
Db 1343 TQCIKQID--HSEKMKF--ITKSSLSDSSVTKSTQW---NSDSHIVTSAFPIDLK 1396
QY 681 QAOLATSLMKISAQACYNESIHELKTR 707
Db 1397 RSPFQNKFSHVQA-----SSYIDFKTR 1419

RESULT 21
US-08-728-470-10
Sequence 10, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

AUTHORS: Gitschler, J.
TITLE: Sequence of the Murine Factor VIII cDNA.
Patent No. 574446
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
US-08-474-503-6

Query Match 2.9%; Score 132.5; DB 1; Length 2319;
Best Local Similarity 20.2%; Pred. No. 0.02;
Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FEDCT-----MESLFPAL-----CAHNSGDDPLVVLNSYCWTFVSKLHITDPKALF 92
DB 775 FKDSTIPKNDMEKIEPQEEIAEKLVQSVSDMLMLQOSH-TPHGFLSDGQEAIV 833
QY 93 K-----EKGDLISQNFRLSFTDCSSKSSPSIIHQKNQOLSIRNNGSMFC 139
DB 834 EAIHDDSPNAIDSNBPSKVTLR-----PESHSEKIVTFPQGLQLRSMKSL--- 883
QY 140 RNHAEGSGGAIADAFSLQINLYFT-----AFEE-----NSSKNGKGA 178
DB 884 ETTIEVWKKLGLQVSSLPNSLMTTLLSDNLKATFEKTDSSGFPDPVHSSSLSTAF 943
QY 179 QAQFSLSRNVSPISFARNADLNGAICCSNLICSGNVMPFTGSAATNGAICCID 238
DB 944 GKRAYSLVGSVPLNASENSD-----SNLDS-----TLWYSQSLSPDNLSIEND 991
QY 239 LNTSEK--GSLSLACNOETLFASNSAKEKGAIV--AKIHVLYRNGPVSEINNSAKIG 293
DB 992 RLKREKRFHGIALTKONTLFKQDVSLMKNTKYNHSTNTEKLTHTSPJSIENSTJLQD 1051
QY 294 AI-----AIOSSGSLSLAGEASYLPONNSQRTSDGLVRAITLLEDAIIS-SLEARNG 347
DB 1052 AILKVNSEIQ---EVTALLHDGTLGKNS------YLRIMHMLNRTTSTK 1095
QY 348 DILF---FDPVQSSSKESPSPSSLOASVTPYPATASPLVIGTSANRSVTFESSERISE 404
DB 1096 DIFIRKDEDPIDODEMTIMFCKMLFS-----ESSNFKKTNGNSL-----NSE 1142
QY 405 EEKTPDNLTSOLOPIELKSGRLVKDRAVLAP---SLSDPQALLIMEAGTSLKTTSSD 461
DB 1143 OEHSPKQVLVLM-----FKKYVKNQSFLEKKNKVTVEODGFTKINIGLKDMAFPHMS 1194
QY 462 LKLTLSLPLHS---LQTEKSV--THAPNLSIOKIFLSN---SGDENFENWELLSKE 512
DB 1195 IFLTTL-S-NVHENGRRHNOEKNOIEIEKALIEEKVYLPVHEATGSKNFKDLILIGTR 1253
QY 513 QN-----NIPLLTSLKQSHLHPDGNLSSHFGVQGMWTFSMKQSDGCHSLIA 560
DB 1254 ONISLEYVHVPLONITSINNTFTVOIHM-----EHFFKRR-----KDETNSEGLV 1301
QY 561 NMTPEKNVYPPAPEROSTVANTLMTVSDMAVOAMINTIAGAVLFGTWSAVSNLEFYA 620
DB 1302 NKTRF-----NVKN-----YPSQKNITQORSKRALGQFRLSTQW---LKTINCS 1342
QY 621 HDSSGKPLDNMHHSLGLFGLSTHSDHSCIAAGOLLKSSDSFTSTETTSYATV 680
DB 1343 TQCIKQID--HSKEMKF--ITKSSLSDSVVIKSTQT--NSSDSHIVKTSAPPIDLK 1396
QY 681 QAOLATSLMKISAQACVNESIHEKLTJK 707
DB 1397 RSPFONKFSHYQA-----SSYIYDKTK 1419

RESULT 18
US-08-670-707A-6
Sequence 6, Application US/08670707A
Patent No. 5859204
GENERAL INFORMATION:

APPLICANT: Lollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSER: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,707A
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95F
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2319 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Mus musculus
PUBLICATION INFORMATION:
AUTHORS: Elder, F.
AUTHORS: Lach, J.
AUTHORS: Gitschler, J.
TITLE: Sequence of the Murine Factor VIII cDNA
Patent No. 5859204
JOURNAL: Genomics
VOLUME: 16
PAGES: 374-379
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 1 TO 2319
US-08-670-707A-6

Query Match 2.9%; Score 132.5; DB 2; Length 2319;
Best Local Similarity 20.2%; Pred. No. 0.02;
Matches 151; Conservative 109; Mismatches 300; Indels 187; Gaps 35;

QY 46 FEDCT-----MESLFPAL-----CAHNSGDDPLVVLNSYCWTFVSKLHITDPKALF 92
DB 775 FKDSTIPKNDMEKIEPQEEIAEKLVQSVSDMLMLQOSH-TPHGFLSDGQEAIV 833
QY 93 K-----EKGDLISQNFRLSFTDCSSKSSPSIIHQKNQOLSIRNNGSMFC 139
DB 834 EAIHDDSPNAIDSNBPSKVTLR-----PESHSEKIVTFPQGLQLRSMKSL--- 883

Db 495 TFWNITANORIVYSSIN-LSNGSLTLMSEGRSGGVEINNDITTGDDTGANLTIYSGG 553

QY 627 PUDMMHRSICLYFGISTHSLDHSFCLAQQL-----GKSS 664

Db 554 WVDHKNISLCAOGNINITAKODIAFEKGNQVITGOGITTSNGNOKFRNNVSLNGTCS 613

QY 665 DSPTSTSTSYATVQALATSL---MKISAQACYNESIHLELTKYRSKESFGSMH- 720

Db 614 GLOFTTKRTKKAITNKFEGTGLNISKGVNISWLPKNEGYD-KFKGRTY-----WNL 665

QY 721 ---SVAQSVCAIPVNSGCLFSSFSIFSKLQCFSGQDCHENSSEIRSFSSSFR 777

Db 666 TSLVSESGEF--NLITDSRGSAGTLTOPYNLNGISFNKD---TTFVNERAVNF- 718

QY 778 NISLPIGI 785

Db 719 DIKAPIGI 726

RESULT 15

US-08-169-927-2

Sequence 2, Application US/08169927

Patent No. 5783441

GENERAL INFORMATION:

APPLICANT: Carl, Mitchell

APPLICANT: Dobson, Michael E.

APPLICANT: Ching, Wei Wei

TITLE OF INVENTION: Gene and Protein Applicable to the

TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Counsel, Naval Medical R & D Command

STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.

CITY: Bethesda

STATE: MD

COUNTRY: USA

ZIP: 20889-5606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/169,927

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/742,128

FILING DATE: 08/09/91

ATTORNEY/AGENT INFORMATION:

NAME: Spevack, A. David

REGISTRATION NUMBER: 24,743

REFERENCE/DOCKET NUMBER: 75,976

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 295-6759

TELEFAX: (301) 295-1022

INFORMATION FOR SEQ. ID NO. 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1612 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-169-927-2

Query Match 3.0%; Score 134.5; DB 1; Length 1612;

Best Local Similarity 19.9%; Pred. No. 0.0064;

Matches 166; Conservative 128; Mismatches 299; Indels 243; Gaps 41;

QY 130 LRNGS---MSTCRNHAEGSGAISADAFSLQHNLYLTAFEENSSKNGAGATQATFSL 185

Db 393 VKNGNTAGVITFMAN-----GALVAST-----DPNIANTINALEAGAV 435

QY 186 SRNVSPISFARNRADLNGAI--CCSNLCSGVNVPFLFTGNSATNGCAICISDLNTSE 243

Db 436 V-ELSGIHIAELRLG-NGSIFKLADGTIVNGVQNAAMNNALAGSI---QLDGA 489

QY 244 -----KGSISLACNDETFASNSAKEKGALIAKIMVLRYPGVFFINSAKIGAI 296

Db 490 IITODIGNGVNAALQIHL-ANDASK-----ILLALDG-ANIT--GAVNGAIH 534

QY 297 IO--SCGSLSTIAGCSVLFQNNNSORTDQGLVRAIYLEKDALISLEARNQDILFFDPI 355

Db 535 FQANGGITKLNTNNINLVNFDLITDITDKITGVYDASSLTNNQTLT---NGSI----- 584

QY 356 VOESSKESPLPSLSQASVSPPTATASPLVIOISANRSVIFSSERLSEBEKPDNLTSQ 415

Db 585 -----GTVAANTKTLAQ---LNGSSKTLINAGDVAINELVLENNQSYQ 625

QY 416 LOOPTEL--KSGRLVKDRVILSAPSLSDOPALLINEAGTSKTS---SDKLATLST 469

Db 626 LNHNTYLTITKTIANAAGQIIVAADPLNTN---TTLADGTNLGSAENPLSTIHFAFKAA 681

QY 470 PLUS-LDPEKSVITHAENLSIQIFLSNGSDENFEKVELLSKQNNIPLUTLSKBSH- 527

Db 682 NADSIILVGVKGVNLINANNITND---ANVCSLHFRSG-----GTSTVGTGGOQGHK 731

QY 528 ---LHLPDGNLSHFGYQGMFTESWMDSCSL--TANNTPKNVYHPEROSTL--VANT 581

Db 732 LNNILUNGTTVK---FLGDTTFNGGFKIEGKSLQISNNYTTDHYESANTGTLEFVNT 788

QY 582 LNMYSQMAVQSNINIAAGATLEFGTNGSAV---SNLFYADSSGKIDNMHHISLG 637

Db 789 -----DPTVYLNKQAG-EGVLKQVIISGPNIVF-----NEIGN-----YG 825

QY 638 YLFCISTHSLDHSFCLAQOLLCKSSDSFTSTSTSYATVQALATSLMKISAQACV 697

Db 826 IVHGIANNSISFENASIGTSLPLPSGTPDLVLTIKSTVGNGTVDNFAPIVYVSGIDSMI 885

QY 698 NESIHELKTKYRSKSGFSGSWHSAVSGEVCASIPVNSGCLFSSFSIFSKLQCFSGT 757

Db 886 NNG--QIICDKKNITALLSDNSITV-----NANTLYSGIIRTKNNQGVTL 931

QY 758 QDGEESSEGEIRSSASFFNISLPIGITFEKKSQKR---TYFFLAIYO----- 806

Db 932 SCGPNNPGFI-----YGLGLENGSPKLQVFTTVDYNNLSIJIANVYTN 977

QY 807 -----DLK-----RDVES-----GPVYLKNA 823

Db 978 DDTLTGTGAGTDFDQAKIILGVSNGNANRFVDSSTSDRSMTIVATQAKGTYTIGNA 1037

QY 824 V-----SWDAPMANLDSRAVMFRLTNORALHRLQTLINVSCVLRGQSHSYSLDGT 874

Db 1038 LVSNIGSLDTPVASV-----RFTGN-----DSGAGLOGNIIYSONIDFGT 1076

RESULT 16

US-08-212-133A-8

Sequence 8, Application US/08212133A

Patent No. 5663060

GENERAL INFORMATION:

APPLICANT: Lollar, John S.

APPLICANT: Runge, Marschall S.

TITLE OF INVENTION: Hybrid Human/Animal Factor VIII

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kilpatrick & Cody

STREET: 100 Peachtree Street

CITY: Atlanta

STATE: Georgia

COUNTRY: US

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk


```

Db 126 NRVTS-----NOISQLGILDSNGQVFL-----INF-----NNTITGKDAII 162
QY 240 NTSEKGSLSLACNOETLFASN-----SAKEKGAIIYAKH--MVLRYNGEVPSPINNSAKTIG 293
Db 163 NTNGFTASTLDISENENIKARFTFEQTKDALKAEIVNHGLITVGKDSVNLIGKVKNEG 222
QY 294 AIAIAGSGSLIAGEGCVLFQNNNSQRTSDQGLVRNATYL-----EKDAI-LSSLERANG 347
Db 223 VLSV-NGCSISLLAG-----OKTITSDIINPTITYSIAAPEVAVNLGDIAPKAG 271
QY 348 DILFFDPIVOESSKESPLPSLSQASVTSPTPATASPLVIQTSANRSVIFSSERLSEEEK 407
Db 272 NI-----NVRATIRNQGKLSADSV 291
QY 408 TPBNULTSOLQOPIELKSGRLVKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTHIAP-----485
Db 340 VIDLSGKKEGETYLGDERGEGKNGIQIAKKTSLKSGSTINVSKEKKGRAIWMGDIALI 399
QY 486 --NLSIOKI-----FLNSGDENFEYENVELLSKEQ-----NNIPLLTSLKESHLHL 530
Db 400 DGNINAGSGSDIAKTGGFVETSGHDLFIKDNAIVDAKEMWLDLDFDNVSIAMETAGRS 455
QY 531 PDGNLSHFPGYQGDWTFSMKDSDEGHSLIANWTPKNVYPAPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTTITNTLESILKKG 494
QY 585 TYSQMA-----VQSMINTIAHGAYLF--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TEVNITANQRIYVNSSIN-LSNGSLTLMSEGRSGGVEINNDITGDDTRCANLTIYSGG 553
QY 627 PIDNMHRSGLYLGISTHSLDHSFCLAGOLL-----GKSS 664
Db 554 WVDVHKNISLGAQGNINITAKODIAFEKGSQVYTGQGITSGNOKGFRFNNSVNLGTGS 613
QY 665 DSIPTSETTSYATVQAQLATSL--MKISAQACYNESIHELKTKYRFSKEGQSMH- 720
Db 614 GLQFTTKRTNKYAITNKEGFLNISGKYVNISMVLPKNESGYD-KFKGRTY-----MNL 665
QY 721 --SVAASGEFCASIPYNSGSLFSSFSIFSKLOGFSGTVDGEEBSGELTSSASSFR 777
Db 666 YSLWVSSGEGE--NLFTIDSKDSQACHTLPYNLNGISFNKD--TFPVVERNAKRVNF- 718
QY 778 NISLPIGI 785
Db 719 DIKAPIGI 726

RESULT 10
US-08-302-832-2
: Sequence 2, Application US/08302832
: Patent No. 5603938
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5603938-Typhale Haemophilus
: NUMBER OF SEQUENCES: 8
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matlare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US Pct/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Belkstrasser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

Query Match 3.0%; Score 134.5; DP 1; Length 1536;
Best Local Similarity 19.4%; Pred. No. 0.0058;
Matches 153; Conservative 123; Mismatches 253; Indels 259; Gaps 38;

QY 120 IHHKNCQLSLRNNGSMSPCRNHAEGSGAISADAFSLQHNLYLFTAFEESSKNGNGAIQ 179
Db 76 VVH--GTATMOYDGNKTIIRNSVD--AIIINMKQFINDENVOPLQEN--NNSAVF 125
QY 180 AQTFSLSRNVSPISFARNRDLNGAICSNLICSGVNPFLFTGNSATNGAICISIDL 239
Db 126 NRVTS-----NOISQLGILDSNGQVFL-----INF-----NNTITGKDAII 162
QY 240 NTSEKGSLSLACNOETLFASN-----SAKEKGAIIYAKH--MVLRYNGEVPSPINNSAKTIG 293
Db 163 NTNGFTASTLDISENENIKARFTFEQTKDALKAEIVNHGLITVGKDSVNLIGKVKNEG 222
QY 294 AIAIAGSGSLIAGEGCVLFQNNNSQRTSDQGLVRNATYL-----EKDAI-LSSLERANG 347
Db 223 VLSV-NGCSISLLAG-----OKTITSDIINPTITYSIAAPEVAVNLGDIAPKAG 271
QY 348 DILFFDPIVOESSKESPLPSLSQASVTSPTPATASPLVIQTSANRSVIFSSERLSEEEK 407
Db 272 NI-----NVRATIRNQGKLSADSV 291
QY 408 TPBNULTSOLQOPIELKSGRLVKDR-----AVLSAPSLSDPPQALLIMEAGTSIKTSS 460
Db 292 SKD-----KSGNIVLSAKEGEAEIGVISAQNOQAKGKLMITGDKVTLTKTGA 339
QY 461 DLKLA-----TLSIPL-HSLDTEKSVTHIAP-----485
Db 340 VIDLSGKKEGETYLGDERGEGKNGIQIAKKTSLKSGSTINVSKEKKGRAIWMGDIALI 399
QY 486 --NLSIOKI-----FLNSGDENFEYENVELLSKEQ-----NNIPLLTSLKESHLHL 530
Db 400 DGNINAGSGSDIAKTGGFVETSGHDLFIKDNAIVDAKEMWLDLDFDNVSIAMETAGRS 455
QY 531 PDGNLSHFPGYQGDWTFSMKDSDEGHSLIANWTPKNVYPAPEROSTLVANTLW-----N 584
Db 456 --NTSEDEYTG-----SGNSAS--TPKR-----NKEKTTITNTLESILKKG 494
QY 585 TYSQMA-----VQSMINTIAHGAYLF--GTWGSAY--SNLFYAHDS-----SGK 626
Db 495 TEVNITANQRIYVNSSIN-LSNGSLTLMSEGRSGGVEINNDITGDDTRCANLTIYSGG 553
QY 627 PIDNMHRSGLYLGISTHSLDHSFCLAGOLL-----GKSS 664

```

QY 563 TRKNVYHPHEROST-----LVANTLNTYSMDQA 591
DB 677 TRFKYISSNSKGLTTOYRSSAGYNGVNGNMFNLKAGAKYFKLKPENMNT-SKPLP 735
QY 592 VQSMINTIHAGAYLEFGTWSAVSNLEFYAHDSGKPID-NMHRSLGYEGISTSLDDH 650
DB 736 IRLANITATG-----GSVFEDYANISGAGELKSEITISNGANFTLSHVRGDD 788
QY 651 SFCLAGOLLKSSDSFIIITETTSYIAVQAOLATSLMKISAQACYNESIHCLKTK-YR 709
DB 789 AFKI-----NKDLTINATNS-----NFSLRQTKDOFYD 816
QY 710 SFSEGEFGSMHVAVSG 726
DB 817 GYARNAINSTYNISILG 833
RESULT 7
US-08-617-697-9
Sequence 9, Application US/08617697
Patent No. 5977336
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5977336-typeable Haemophilus
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-9
Query Match 3.2% Score 146 DB 2; Length 1599;
Best Local Similarity 18.5% Pred No. 0.00051;
Matches 153; Conservative 143; Mismatches 312; Indels 218; Gaps 39;
QY 73 NSYCMFVSKLHTDPREALFKENG-----DISLONRFLSTDCSSKSPSJIHOKNG 126
DB 143 NGCVFLINPGITIGKDAINTNGFTASTLDSINENIKANFTLEDJTKKALAEI-VNHG 201

QY 127 QLSLRNNGSMSTCRNHAESGGAISADAFSLQHNILFTAFEESSKNGCAIO---AQTF 183
DB 202 LITVCKDGSVNL-----CGKVK-----NEGIVSVNGSISLLAGOKI 239
QY 184 SISRANSP-----ISPARRADLNGAICCSNLCSGYNVPLFFPGNSNTNGAICICSD 238
DB 240 TISDIITPITITSIAPEDEA-INLGDIATK-----GGIN-----VRA 277
QY 239 LMTSEKSLSLACNOETLFPASNSAKEGGAIV-AKHVILRYNGPVSPFNNSAK-----290
DB 278 ATIRNKGRKLS-----ADSVSKDKSGNIYLSAKEGAELIGVYISAQNOQAKGKGLMI 328
QY 291 IGAATAIOSGSLSLIAGSGSVLPONNSORTSDQGLVRRAITLEDALISLEARN-----346
DB 329 TGDKVTYLTGAVIDLSGKEGETYLOGDER-----GEGKNGIOLAKKTTEKSTINVSOK 384
QY 347 -----GDILFPDPIVQESSKSPSSLOASVTSPTPATASPLVIOTSA-----391
DB 385 EKGRALVWGDIALIDGNI-----NAQKDKIKTG-----FVETSGHYLSD 427
QY 392 NNSVIFSSERI,SEEEKTPDNLISQLOPIELKSGRLVLDRAVLASPSISODPOALLIME 451
DB 428 DNAIVTKEMLLD-----PENVTIAPSASRVELG-----ADRSHSASEVIK-----VTLKK 474
QY 452 AGSLKTSDDLATLSTPLH--SLDTEKSVTHAPNLSIOK---TFLNSGDENFYENV 506
DB 475 NNTSLTPTLNTTISNLKSAHAVNITARRKLTIVNS-SISEKSHLILHSEOGG--QCV 531
QY 507 EL---LSKQNNPIPLTLTSKEQSHLIPDG---NLSHFYOGDPTFGMKSDSECHSLI 559
DB 532 QIDKDTSEGNLTISGQWVDYHKNTITLGSGLNITTK---EGDAFEDKSGRNNLTIT 588
QY 560 ANWTPKNVYHPHEROSTLVANTLMN--TYSDM---QAVQSMINTIAHGAYLEFGTWS 612
DB 589 AOGTITSGNSNGRFNNVSLNSIGKLSFTDSREDGRRTKGNISMKFGDTLINISOT---645
QY 613 AVSNLEFYAIDSSGKPIDNMHRSGLYLF-GISHSLDHSFCLAAQLLCKSDSTITST 671
DB 646 -----VDISMKAPKVSFWFRDCKRTYNNVTTLVN-----TSGSKRNLSTI 684
QY 672 ETSYIATVQAOLATSLMKISAQACYNESIHCLKTKYRSFSEKSGFSGMHSVAVSGEVCAS 731
DB 685 DSTGSGST-----GPSIRNAELGITFNKATFNIAAGSTANFSIKAS 726
QY 732 I-PIVNGSGLEFSFISFKLOGFSGTODCFEESGSEIRSFASSFRTNISLPITFEKK 790
DB 727 IMPEKSN-----ANVALFNEIDISVG-----GGSVNFKLASSNITQTP-GVII--K 770
QY 791 SQKTRTYIFLAGYIODLKRDESQVVLKKNVSDADAMANDSR 836
DB 771 SQN-----FNVSGSTLNLKAEGSTETAFSIENTDLNLTATGNTITIR 812
RESULT 8
US-08-728-470-9
Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS-VG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-469-880-4

Query Match 3.3%; Score 151.5; DB 2; Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

73 NSCYMFVSKLHTDPEALFKEKG-----DLSTQNFRLSTFDCSSKSS-PSIIHQ-- 123
143 NGCVFLINNGITIGDALINTNGFLASTLDSNENIKARNFTFEOTKALAEIVNHL 202
124 ---KNGQSL-----RNGSMSFECRNHAEGSGAIS-----ADAFSLQHNLYLFA 165
203 ITVGKGSVNLIGKAKVNEGVISV-----NGGSISLAGOKITISDILNPIIVYSIAA 255
166 FEENS-----SKG-----NGCAIQAFPSLSRNVSPIFANKKADI,NGCAIC 207
256 PENEAVNLCDIFAKGNINVRATTIRNOKLSADSVSKRSGIVLSAKGEAEIIGVIS 315
208 CSNLICSGVNPFLFPGNSAT--NGCATCCISDINTSEKGSLSLACNOFTLFAASNAKE-K 265
316 AQNOQAKG--GKIMITGDKVTLTKGAVI---DLSCKEGG-----EYLGSDERGECK 362
266 GGAIVAKHMYLRNGPVSFINSKAKIGAIATQSGSLSLAGEGSLVFQNNQSRTSDOG 325
363 NGIQLAKKTSLEKG---STINVSCKEKGGAIVWC---DIALIDGININAGSGDIAKTG 416
326 LVNNA---IYLEKDALISLEARNCDILFPDPIVORSSSKESPLPSSLSQASVSPTRPFA 382
417 FVETSGHDLFEIKDIAVDAKEW---LLDPDNY---SINAEPL----- 453
383 SPLVIOTSANRSVIFSSERSEFEKTPDNLTSOLOQPIELK-----SGRL----- 427
454 -----FNNTCINDEFPICTGEASPKKNSLKTLLTNTTISNYLKNAWTMN 499
428 VLKDRAYLAPSLSDPPALLM---EAGTSKTSDDLKATLSTPLHS---LDTEKSV 480
500 ITASRKLTIVSSINIGSNHLLHSKQQRGGQYQJODLITSKGNNLTIVSCGVADVHKNT 559
481 TIHAPNLSIOKIFLNSGDNFYE---NVELLSK-----KNNIPL-----L 519
560 TLDOGFINTAASVALEGNNKARDAANAKIVAGVTTITTEGCKDRFANVSLNGTGKGL 619
520 TLSEKOSHLHPDGNLSSHGQYQDWTFSS-----WKSDSGHSLI-----ANW 562
620 NIISSVANL---THNLSCTINISGNITITNTTTRKNTSYWQTSIDSHSNVSAALNLETCANF 676
563 TPKNYVPHEROST-----LVANTLMNTYSQMOA 591

677 TFIKYSNSKGLTYQRRSSAGVNEGVNGMSTNLEKAGAVNRKLRKNEKMT-SKPLP 735
592 VQSMINTIAHGCAYLFGWCSAVENLFYAHDSCKPID-NMHHRSGLYLCISTHSLDDH 650
736 IRLPLANTATG-----GSVFIDIVNHSRGAEKMSFINISNGANFLNSHVAGDD 788
651 SECLAAGLCKSSDSFSTSTETSYIAFVQALATSLMKISAQACINESHEIKTK-YR 709
789 AFKI-----NKDLTINATNS-----NFSLRQTKDQFYD 816
710 SFSKEGFGSHSVAVSG 726
817 GYARNAINSTYNISILG 833

RESULT 5
US-08-728-470-4
Sequence 4, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Matlare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-728-470-4

Query Match 3.3%; Score 151.5; DB 2; Length 1477;
Best Local Similarity 18.9%; Pred. No. 0.00013;
Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

73 NSCYMFVSKLHTDPEALFKEKG-----DLSTQNFRLSTFDCSSKSS-PSIIHQ-- 123

```

Db 143 NCOVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNTFEOTKDKALAEIVNHL 202
QY 124 -----KNGQLSL-----RNNGSMFCRNHAEGSGGALS-----ADAFSLQHNLYFTA 165
Db 203 ITVGKGSVNLIGKVKYKNEGIVISV-----NGGSISLLAGOKITTSIDINPTITYSIAA 255
QY 166 FEENS-----SKG-----NGCAIOAOTFSLSRNVPISFARRADLNGAIC 207
Db 256 PENEAVNLQDIFAKGNINVRATINQCKLSDVSKDSKSGNIVLSAKGEAEIGVIS 315
QY 208 CSNLICSGVNPLEFTGNSAT-NGAICISDLNLTSEKSLSLACNOETLFAANSKAE-K 265
Db 316 AQNOQAKG--GKLMITGDKVFLTKTGAVI---DLSGKEG-----ETVLGDEREGEK 362
QY 266 GGAIVAKHMYLRNGPVSPFINNSAKIGGAIAIOSGSLSLAGEGSLFQNNQSORTSDG 325
Db 363 NGIOLAKKTSLEK-----STINVSKEKGGFAIVMG---DIALIDGINAGSGDIAKTGG 416
QY 326 LVRNA---IYLEKDAIILSLFARNCDLFPDPPIVOESSKESPLPSIAQSVISPTPATA 382
Db 417 FVETSGHDLFIKNAIYDAKEM---LDFEDNV---SINMEDPL-----453
QY 383 SPLVIOTANSRVSIFSSERLSEKTPDNLTSQLQPIELK-----SGRL-----427
Db 454 -----FNNTGINDERPTGTGEASDPKKNSLKTTLTNTTISNYLKNAMTMN 499
QY 428 VLKDRAVLSPSLQDPOALLIM---EAGTSLKTSDDLKATLSIPLHS---LDTEKSV 480
Db 500 ITASRKLTVNSINIGSNHLSHSGQGGVQIDGDTISKGNLTITYSGGWVDVHKNI 559
QY 481 TIHAPNLSIOKIFLNSGDENFYE---NVELLSK-----EQNNIPL-----L 519
Db 560 TLDGFLNITAAVAFEQGNKARDAANAKIVAGTVITTEGCKDFRANNVSLNGTGKL 619
QY 520 TLSKQSHLHPDGNLSHFGYGDWTF-----WKDSDEGHSLL-----AMN 562
Db 620 NIISSVNLT---TNHLSGTINISGNITITNOTTRKNTSYWQTSKSHMNVSALENTGANKF 676
QY 563 TPKNYPHPEROST-----LVANTLMNTYSMDQA 591
Db 677 TFIKYSNSKGLTTOYRSSAGVNFNGNMFNLKEGAKVNFKLKNENMNT-SKRLP 735
QY 592 VQSMINTIAHGAYLFGTWSAVSNLFYAHDSGKPID-NMHRSLQYLFGISHSLLDDH 650
Db 736 IRLANITATAG-----GSVFEDIYANHSRGAEELKMEINISNANFTLNSHVAGDD 788
QY 651 SFCLAAQQLCKSDSRITSTSTSYIATVQAOLATSLMKISQAQVNESIHELKTK-YR 709
Db 789 AFKI-----NKDLTINATNS-----NFSLRQTKDUPYD 816
QY 710 SFSKEGFGSMHSAVSG 726
Db 817 GYARNAINSTINISILG 833

```

```

RESULT 2
US-08-302-832-4
Sequence 4 Application US/08302832
Patent No. 5603938
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mataré, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302.832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry M
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

```

Query Match 3.3%; Score 151.5; DB 1; Length 1477;

Best Local Similarity 18.9%; Pred. No. 0.00013;

Matches 151; Conservative 123; Mismatches 274; Indels 249; Gaps 36;

```

QY 73 NSYCWPFYKLIITDPKEALFEKKG-----DLSIONFRFLSFDSCSKRESS-PSIIHQ-- 123
Db 143 NCOVFLINPGITIGKDAIINTNGFTASTLIDISNENIKARNTFEOTKDKALAEIVNHL 202
QY 124 -----KNGQLSL-----RNNGSMFCRNHAEGSGGALS-----ADAFSLQHNLYFTA 165
Db 203 ITVGKGSVNLIGKVKYKNEGIVISV-----NGGSISLLAGOKITTSIDINPTITYSIAA 255
QY 166 FEENS-----SKG-----NGCAIOAOTFSLSRNVPISFARRADLNGAIC 207
Db 256 PENEAVNLQDIFAKGNINVRATINQCKLSDVSKDSKSGNIVLSAKGEAEIGVIS 315
QY 208 CSNLICSGVNPLEFTGNSAT-NGAICISDLNLTSEKSLSLACNOETLFAANSKAE-K 265
Db 316 AQNOQAKG--GKLMITGDKVFLTKTGAVI---DLSGKEG-----ETVLGDEREGEK 362
QY 266 GGAIVAKHMYLRNGPVSPFINNSAKIGGAIAIOSGSLSLAGEGSLFQNNQSORTSDG 325
Db 363 NGIOLAKKTSLEK-----STINVSKEKGGFAIVMG---DIALIDGINAGSGDIAKTGG 416
QY 326 LVRNA---IYLEKDAIILSLFARNCDLFPDPPIVOESSKESPLPSIAQSVISPTPATA 382
Db 417 FVETSGHDLFIKNAIYDAKEM---LDFEDNV---SINMEDPL-----453
QY 383 SPLVIOTANSRVSIFSSERLSEKTPDNLTSQLQPIELK-----SGRL-----427
Db 454 -----FNNTGINDERPTGTGEASDPKKNSLKTTLTNTTISNYLKNAMTMN 499
QY 428 VLKDRAVLSPSLQDPOALLIM---EAGTSLKTSDDLKATLSIPLHS---LDTEKSV 480
Db 500 ITASRKLTVNSINIGSNHLSHSGQGGVQIDGDTISKGNLTITYSGGWVDVHKNI 559
QY 481 TIHAPNLSIOKIFLNSGDENFYE---NVELLSK-----EQNNIPL-----L 519
Db 560 TLDGFLNITAAVAFEQGNKARDAANAKIVAGTVITTEGCKDFRANNVSLNGTGKL 619
QY 520 TLSKQSHLHPDGNLSHFGYGDWTF-----WKDSDEGHSLL-----AMN 562
Db 620 NIISSVNLT---TNHLSGTINISGNITITNOTTRKNTSYWQTSKSHMNVSALENTGANKF 676

```


Db 329 ICAYVNIQGGCPLEFSNNRREKGCALFICHSVDDPAKQTSITLTIILASEGILAFQGMKLN 388
QY 321 TSDQGLVNAIYLEKDALISLSLEARNGD-ILFPDPIVOESSKSEPLSSLOASVTSPT 379
Db 389 TKPG--IENATTEVGCGLIVLSAOGSGRLVFYDPIYH-----SLPTTSP 431
QY 380 AYASPLVYQTSANRNVIFSSERLSEEE-KTPDNLTSQLOPIELKSGRLVKRAVLSAP 438
Db 432 SNKDITIANAGASGVSTFSKGLSTELLIPANTITLLGTGKIASGELKITDAAVNVNL 491
QY 439 SLSDPPALLIMKACTSLKTSISDLKLAITLSIRLMSLDTREKSVTIHAPNLISQIKFLS--- 495
Db 492 GFATQSGQLLIGSG-----GTGLAPPTGAPAAVDDTIGKLAIDPSPFLKRDVFASAV 545
QY 496 NSGDENFEYENVELLSKEONNIPILLTSKEOSHLL-----PDGNLS--SH 538
Db 546 NAGTKNVTITGALVIDEHDYTDLIDWVSLQSPVALPIAPVFGATVTKTGIFDGLIAPSH 605
QY 539 FCGYGDWTFSPK-----DSDECHSLIANWTP-----KNYVHPRKQSTLV 578
Db 606 YGYGCKMSYTSRPLIPAPDGFPGGSPSANTLVAWNSDTLVRSYIIDPERYGEIV 665
QY 579 ANTLNNTSMDQAOVSMN---TAHGCAYILFGWGSANSLPFAHDSCKPIONMHRS 635
Db 666 SNSLWISFLGNQAFSDILQDVLLIDHPGLST---TAKALGAVEHNPFGCHGFGSGRY 720
QY 636 LGYLFGISTHSLDHSFCLAGOLLGK-SSDFTSTETSTYATVQOLATSLMK--- 690
Db 721 CGYQALISMNTDHTTLOLSFGOLXGKTANAPYDSRCEQYILSLFQOPFLVYQKSLAL 780
QY 691 ISAQACWESIHELKTKYKSPSK--ECFGSMHSAVSGEVCASLPVSNQGLSPSPSIF 748
Db 781 ISMKAYGYSKNHLNTTYLTPDKAPKSGQWNNHNSYVLLISAEHP-----FLNMCLL 832
QY 749 SK-----LOGF-----SGTQDGFRESSGELISFSASSFRRNISLPIGITFE-----KK 790
Db 832 TRPLAQANDLSGFIISAEFLGQMSQFTEQDLOHSFSGCKGVNWSLPIGSSQWFTPPKK 892
QY 791 SOKTETYYEFLGAYTODEKRDVESGPLYLKNVAVSWDAPMANL----- 833
Db 893 APSTLTITKL---ATKPDYIRVNPRIJYTVVNSQESTISICANLRRHGLFVQIHIDVVDLITE 949
QY 834 DSRA---YMFRLTNORALRLQOT 853
Db 950 DTQAFNLNTFGCKNGFTNHRVST 972
RESULT 25
H72074
polymorphic membrane protein E/F family CP0286 [imported] - Chlamydomonada pneumoniae (strain
C:Species: Chlamydomonada pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: H72074; E81593
R:Kakman, S.; Mitchell, W.; Marathe, R.; Lammle, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: H72074
A:Molecule type: DNA
A:Residues: 1-938 <ARN>
A:Cross-references: GB:AE001631; GB:AE001363; NID:g4376750; PIDN:AD18608.1; PID:g4376757
A:Experimental source: strain CML029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
N.C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
J. Nat. Acad. Sci. 96, 1397-1406, 1999
A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: E81593
A:Molecule type: DNA
A:Residues: 1-938 <REA>
A:Cross-references: GB:AE002190; GB:AE002161; NID:g7189209; PIDN:AF38143.1; PID:g7189222
A:Experimental source: strain AR39, HL cells

C;Genetics:
A;Gene: pmp_15; CP0286

Query Match	11.9%	Score 540.5;	DB 2;	Length 938;
Best Local Similarity	25.2%	Pred. No. 3.9e-26;		
Matches 229;	Conservative 122;	Mismatches 357;	Indels 201;	Gaps 38;

```

QY 124 KNGQSLFNNNSMSPFCRRHAGCSGALISAD---AFSLQHNLFY-----164
  || : : : : : ||||| |||||
Db 77 KNLLIS-ENIGVFEKKVNCVSGALYIAAONCTISKQNAFTTNVSDNPATAGSL 135
QY 165 -----AFENSSKMGCAIOAQF-FSLSRNYSPISFANRA-----DL 201
  || : : : : : ||||| |||||
Db 136 GGALFALNCSTTNNGQSTFYDNLALKGALYIETNLSIDNKGPIIIKONRALNSDSL 195
QY 202 NGCAICCSNLLICSGVNPFLFTGNSATNGALICCSIDLNTSEKSL-----247
  || : : : : : ||||| |||||
Db 196 GGGIYSGNSLMEGNSGAIQITSSNMSGCGGIFMQTLTISNNKLLIENSNAFANNYG 255
QY 248 -----SLACNOE-TLPASNSAKKEGGALYIAKHVLRINGPVSTNNSKI 291
  || : : : : : ||||| |||||
Db 256 SNFNFGGGGLTTTCTTCTILNREGVLFNNQOSQNGAIIHAKSIIEKKGPPVFLNNTATR 315
QY 292 GGAI-----AIOGGGSLSLAGEGSLVLPONNSORTDQGL---VNRATYLEKDALISSLEA 344
  || : : : : : ||||| |||||
Db 316 GGALLNLSAGSGNSGSLISADNGDIFNNNT--ASKAHNLNPPYRNALHSTNNML-QIGA 372
QY 345 RNG-DILEFDEIYVESSSKESPLPSSLOASVSTPTPATASPLV--IQTSAHSVIFSSER 401
  || : : : : : ||||| |||||
Db 373 RPYCVHLYEYDPI-----EHELPSF-----DILENFEGHGTGLVPSGEH 412
QY 402 LSEERKTNDNLTSOLOPTELKSGRLVKDAVLASPSLSDPPQALL-----IMEAGT- 454
  || : : : : : ||||| |||||
Db 413 VHQNTDEMNFEFYSYLRNLSRCKOGVLAVEDAGLACYKFPORGGLLGGAGVITTAGTI 472
QY 455 -----SLKTS--SDKLATLSIPLHSL-----DREKSVTGHAP 485
  || : : : : : ||||| |||||
Db 473 PTPSSTPTTVGSTITLNIATIDPLRISFOQARKIMVITPTKGTSTYEDSNPTITISG- 531
QY 486 NLSIOKIFLNSGDFENYENWELLSKEONNIPRLTL-----SKEOSHILDPDNLSSH 539
  || : : : : : ||||| |||||
Db 532 -----TLLRMSNNDPYDSL-LSHLEKVPILYIVDAQAQINSQDLSTJNSGEHY 585
QY 540 GYQGWMTSMKD-----SDEGSHLI-ANMTPKKNVPIPERQSTLVANTLNTY 586
  || : : : : : ||||| |||||
Db 586 GYGWISYVWVETTTITNPTSLGANTKHKLLYANMSPLGRRPHERGERTIALM-- 642
QY 587 SOMQAVQSMIMNIAHGAVYLEGT---GSVSV---NLGYAHQSSKPIIDNNHNRSLG 637
  || : : : : : ||||| |||||
Db 643 -----QSAIYTLA--GLHSLSSWDEEGGHAASLOGIGLLVHQDKNG--FKGFRSHMTG 692
QY 638 YLFGLSTSLDDHSPCLAGOLLEKSSD-SFTSTETTSYA--TVQAOALTSIMKISSAQ 694
  || : : : : : ||||| |||||
Db 693 YSATTEANSSQSPNLSLOFAQFFKSAKHEHSQNSTSHNHYSQMCIENTLKRMIARLSVS 752
QY 695 ACYNASHIELKTKYNSFSK-EGFGMSHVAVNGEV-CASIRIVNSGSLJESSFSIFSLO 752
  || : : : : : ||||| |||||
Db 753 LAYMTSEHTITMYOGLLEGNSSQGFNNHTLGAALSCYFLP-QPRGESL-QIYEPITAL 809
QY 753 GFSHQDGGFEESSEGEIRSFSA-SSEFRNLSLPIGTFE-KKSOQRTYUUVFLGAYIODKR 810
  || : : : : : ||||| |||||
Db 810 AIRGMALAFQESGBIHAREFSLHRLPTDVSLEPVGIASKNNHNRVPLVLTETISVSTIR 869
QY 811 DVESGPVYLKNAVSWDAPMANLDSRAYMFLTNORALHRLQTL-LNVSCVLRQSHSYS 869
  || : : : : : ||||| |||||
Db 870 QDPELHKSULLISOGWITQAPVPYTNALGIKVKNTMQVFPKVTISLDSVADISSSTISHY 929
QY 870 LDLGTTIRF 878
  || : : : : : ||||| |||||
Db 930 LNVASRMRF 938
  || : : : : : ||||| |||||

```

Oy 376 SPTPATASLVYIOTANRNVISSEERLSEEEKTPDNLTSOLOQIETLKSGVLVCKRAVL 435
 Db 419 -----GTLFSSYYIPDITSRDDEIFSHFRNHICLYNGTLAEDRAEW 461
 Oy 436 SAPLSODPOALLI-----MEACTSLKTSDDLATLTLSPHSL----- 474
 Db 462 KYRKPDQFCGTLRLGSRVFEFTDEOSSSSVGVININNLAIPLPILGNRAVAKMLIR 521
 Oy 475 -----DPEKSVTHAHPNLSTOKIFLNSGDENF--YENVELSKEDONNIPLL-TL 521
 Db 522 PTGSSAPYSEDNPIINLSGP-----LSLDDENLDPDYDAD-LAQPIAEYPLLYL 572
 Oy 522 SKEOSHLH----LPDG-NLSSHEGYGDMTFMSKD-----SDEG-----HSLIANT 563
 Db 573 DVTAKHINDNFYPECLNTTYQHGYGQWSPWYLETTTSPTSEEDIVNTLHKOLYGDWT 632
 Oy 564 PKNYVPHPEROSTFLVANTLTNMTYSDM-----QAVQSNININTIAHGAYI.PGTWGSVANSU. 617
 Db 633 PTGYKKNPENKGDIALSAFQSFHNLFATLRGYTOQGGIAPTASGEA-----TR 681
 Oy 618 FYAHDSGKRPIDNMHHKSLCYLFCISTHSLIDHSFCLAGOLLKSSDSFITTSTTYSI 677
 Db 682 LEVHONSNNDAKQFHEEATQYSLGTTSTNTASNHSFCVNFSQLFSLNLYESH--SDNSVASHT 740
 Oy 678 ATVOAQLATSLMK-----ISAQACYNESIHELKTYRSPSKSGFGSGHISVAEGVCASIP 733
 Db 741 TTVVALDINNPWLOERSTSTASIALSYSNHIIKASGYSEKJOTEGCYSTTLGCAALSCISL 800
 Oy 734 IVSNGSGLFESSFISFKLQCFSGQDGFESSGCELRPSA--SSFNNISLPIGITEPKSQ 792
 Db 801 LQMRSRPL--HFFPFIQAIIVRSNQTAFOESGDKARKRSVHKPLXNLNLYPLGIQSAMESK 858
 Oy 793 -KRTYYVFLGAYIIDLKRDVEGSPVYLLKNAVSDADPMANLDSAYVFRILTNRALH-R 850
 Db 859 FRLPYWNIELEAVOPYLYOONPEINVSLESSGSSWLLSGTTLARNAAIFKGRNOIFIFPK 918
 Oy 851 LQTLNVSCVLRGOSHSXSLDGTTFR 878
 Db 919 LSVFLDYQGSVSSSTTYHYLHAGTTFKF 946
 RESULT 22
 D72067
 polymorphic membrane protein A family CP0213 [imported] - Chlamydothilla pneumoniae (str. C3) [species: Chlamydothilla pneumoniae]
 C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C.Accession: D72067 | DB1601
 R.Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999
 A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: D72067
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-947 <ARN>
 A:Cross-references: GB:AE001638; GB:AE001363; NID:94376819; PIDN:AAID18679.1; PID:94376887
 A:Experimental source: Strain CWL029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255
 A:Accession: DB1601
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-947 <REA>
 A:Cross-references: GB:AE002182; GB:AE002161; NID:97189140; PIDN:AAF38083.1; PID:97189140
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: pmp_19; CP0213

Best Local Similarity 25.38; Pred. No. 9.6e-28;
Matches 252; Conservative 122; Mismatches 376; Indels 246; Gaps 41

54 LPPALCAHNSODDPLVYICNS-----YCMVSKLHTIDPEALFEKKGDL5 99
Db 27 LPLSGIHSGEDLEFLLTRSSSPTKTTYSLRDLFYCDPAGN-SHKPCAALNLKGDLE 85
QY 100 ION-----FRLESTDCSSKESSPSIIHOKNCOLSR 131
Db 86 FINSPTLALFTKKNJHLGARGAGLFSESVTFEFGKLSHLENNESSGCVL-TTSGDLSFI 144
QY 132 NNGSMFCCNHNAGSGAL-----SADARSLOHN---YLF-----TAFEENSSKNGGALQ 179
Db 145 NNTSYLCNNISISGPGGALLLOGRKSKALFRRNRGTLIFLKNKAYNODESHPIGGAVS 204
QY 180 AOTYSLSRNVSPISFA-----RARRADLNC-----GAICCSNLICGANNPLFETCN 225
Db 205 ----SISPG-SPITFDNQNELLFQENEGELGAIYMDGCAIFEN-----NQOTSFEN 254
QY 226 SATNGAICCIDSLNTSEKGSULACN-----OETLFASNAKEKGAIYAKHMYL--- 277
Db 255 KASPGGAV-----YSRYCNLYSOMDWLFTFNAAKVGAHADVHILROCK 301
QY 278 -----YNGPSPFINNSA-KIGGAIIQSGGS-LSIAGE 309
Db 302 GSTVFEENSTAGGALVAVNACVDCINAGPREFINNSALGLNGAIYMQATISLRHAAQ 361
QY 310 GSVLFPQNSORT-----SDOGLVKNALYLEKDAILSLSEARNGD-ILFDPDIYOESSS 361
Db 362 GDIEFGCNKXRSQFHSIHINSTNFTNNALITQCAPREFSLANEGRICFYDPIISATEN 421
QY 362 KESPLBSSLOASVTPRPATASPLVIO-----TSANSVIFSSRSLSEDEKTPD-NLTSOL 416
Db 422 YNS-----LYNHORLLEGAGAVIFSGARLSPEHKKEKNKNTSII 461
QY 417 QOPTELSGRGLVKDRAVLSAPSLSDOPOLLMEAGTSLKT-----SSDLKLATLSIP 470
Db 462 NOPPYRLCSGLVSLIEGGAIIIVRSFYDE-GGLLALGCSKTLTGOKNSEKOKIYITMUGN 520
QY 471 LHSIDTEKSVTTHAPNLISIOKIFELNSG-----DENFYEYENLLSEONNIPLLTSLK 523
Db 521 LENIDSSDPAEIRA-----TEKASIEISGVPRVYGHIESFEYENHEVASKPYTTSIIISAKK 576
QY 524 EOSHLHPDGNLSS-----HEFGQGDWMTSW--KQSDDEHSHILANMTPK-NVYPPH 571
Db 577 LVYAPSPHEKDIOULIAESEYMOYGQGSWEFSSWSPNDKEKKTIIIASVTPRGEFSLD 636
QY 572 EROSTIVANTLMTYSDMQAVQSMINTIAGGAYLFCT-----WQSAV----- 614
Db 637 KRKRSFPTLMTSFGSLNIASNTVN-----NNTLNNSEVIPILOHLCVFCGPPYQVQMEON 691
QY 615 ----SNLFYAHDSGKPLDNNMHRSLGVLFGISTHSLDHSHFLCPLAAGOLLKSSDSFIT 669
Db 692 PKQSSNNMLVQHAG-----HNVGARIFPFSFT-----ILSALTOLOFSSSSQONVA 737
QY 670 -STETTSYIVYQAQALATLSMKISIAQACVMEHIELK--TKRYSFSKEGSHMSYAVSG 726
Db 738 DKSHAQILIGVSLNKSWMQALSLSSFSYVEDSOVMKHVPYKQTSR--GSMRWYWSG 794
QY 727 EVCASIPVNSGSL-FSSSIFESKLOGFSGTODGFESSGELIRFSASSFRNLSIPGI 785
Db 795 SVGMS----YAPKRIIRLKMTPFVDIOTKLVQNPFEVETGDPRIYFSSSEMTNLSDIGI 851
QY 786 TFEKK--SQKTRITYYELGAIYIDLKADVESGPVYLLKNAYSDAPMANLDSRYMYRLT 843
Db 852 ALENRFISGRSLSLEQVSTYICKLRRNPPOSSASLVNLHYTMDIOGVPLGKALENLITLN 911
QY 844 NORALHRLQTLNLSVYL-RGSHSYSLDLCTTYRF 878
Db 912 STIKKYIVTAVMGISSTQRESNLSAANAAGLSLSF 947

Query Match	12.5%; Score 566; DB 2; Length 947;
-------------	-------------------------------------

RESULT 23

```

Db      182 SVSFYONAAITFGGAIHSSGPLOIAVNOAEIRFQONTAKNGSGALSDSDIDIDQAAAYVL 241
      221 FFGNSAT----NGAICCSIDLTSEKCSL--SLACNOETLFASNAKKGCAIVAKHHV 275
      242 FRENELTTAIGKGAIVCCLPITSGSSTPVPIYFSDNKQIVFERNISIMKGATIAKKLS 301
      276 LRYNGVSTFIN-----NSAKIGGAIAIOSGSLSLAGESVLFQNNNSORTSDQGLVRNA 330
      302 ISSGCTTLFINNTSVANSQNLGALAIIDTGGELSLSAEKGTIFQGN--KTLSPFL--NG 357
      331 IYLEKDAIILSLAENG--DILFDPPIVQESSKESPLPSLQASVTSPPPATASPLVIQT 369
      358 ILLQNAKFLKLOARNGSYIEFYDITSEAD-----GSTQLNIDPK-----NK 402
      390 SANRSVIFSSERLSEEEKTPDNLTSOLOPTELKSGRIYKIDRAVLSAPSLSDPPALLI 449
      403 EYTGTLTFSGEKLAND--PRDKSTIPQVNVNSAGLYLKEGAETVSKETSPSHLV 460
      450 MEAGTSLKTS--SDLKATLSLPLSLDTEKSVTHAPNLSIQKIFLNSGD-----ENFY 503
      461 LDGTLTLAKEDIALITGLAIDTSLSSSTAAVIKANTANKQISVYDIELSPGIMAY 520
      504 ENVELLSKEONNIPLLTSLK--EQSHLHPDGN--LSSHGOGDMTFESKSDSDE--GIIS 557
      521 EDLRM--RNSOTFPLLEPGAGSVTVTAGDPLVSPHYFGQNMKLAWTGTGNKVGEE 578
      558 LIAMTPKNVPPPEROSTLVANTLMTYSDMAVOSMINTIAHGAVLFCITMGSAVSNL 617
      579 F---WQKINKPRPEKGNLVPNILMGNAVDVRSLMQVQETHASSLOTDRGLIDGIGNF 635
      618 FYAHDSGKRPIDMMHRSILYFGISTHSLDDHSCFLAAQLLGSKSDSFTSTETTSYI 677
      636 F--HVSASEMINIRYHNSCQVLSVNNELTPKHYTSMAPQLSRQDVAVSNNEYMYL 693
      678 ATVOAOLATSLMKISAQA-----CYNESIHELKTYRS 710
      694 GSYLYQYTTSLAGNIFRYASNPVNVNGILSRFLQNLMTFHLFCAYGIAHTNMOKTIDYAN 753
      711 FSKRGQSMHSAVSGCVASIPVSMGSG--LFSFSLFSLKLOCFSTQDQPFESSQEI 769
      754 FPMVK--NSMNNNCWALECGSMPLVFENGRLOGAIPFKLQLVVAYQDDFKTTIDGR 812
      770 SFASFSFRNLSLPIGTFEERKSQKTRTYFFLAGYIIDLKRDVESGPNVLKNAVSADP 829
      813 RFSNGSLTISVPLGICFEXLALXQDVLDPFSFYPIDIRKPOSCAALVSDSMILVP 872
      830 MANDSRAVYFRLTNOALHRLQTL--NVSCVLRGQSHSYSLDCTTYRF 878
      873 AAHVSRHAFVSGGTGRYHFNDYTELLCRGSIETC--RPHANNYNINCSKFRF 922

RESULT 19
E72130
polymorphic membrane protein G family CP0761 (imported) - Chlamydia pneumoniae (stra
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: E72130; G81541
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: E72130
A:Molecule type: DNA
A:Residues: 1,841 <RAN>
A:Cross-references: GB:AE001586; GB:AE001363; NID:94376263; PIDN:AD18172.1; PID:9437627
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20130255
A:Accession: G81541

```

```

A:Molecule type: DNA
A:Residues: 1,841 <RAN>
A:Cross-references: GB:AE002235; GB:AE002161; NID:97189672; PIDN:AAF38561.1; PID:9718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: pmp_2; CP0761

```

Query Match 15.6%; Score 708.5; DB 2; Length 841;

Best local similarity 28.5%; Pred. No. 7.8e-37;

Matches 241; Conservative 137; Mismatches 348; Indels 121; Gaps 31;

```

      107 SFTDCSSKSS-----PSIIHQ-----KNGQLSRNNGSNSEFCRNHAEGSGAISAD 153
      41 STFSFSSKTSASADGTNYVFKDSVLENPVKTGETO-----STSCPKN-----DAAGD 89
      154 AFSLOHNYLFTAEENSSKNGCAIQAO-----FSLSRVNSPISTFARNAD--LNG-GAI 206
      90 LNFGLGGFSTFSTNIDATTAAGAISSSEANKVTLIS--GFSALSTLSPSTVNTGLAI 148
      207 -CCSNLISGVNPLPFTGNSATNGAICISDNLNTEKSGSLSLACNOETLFASNAKKE 265
      149 NVKGNLSLIDNOKVLLQDNFSTGDCGALNC-----AGSLKIANNKSLSPGNSSSTR 200
      266 GGAIVAKHMYLRYNGPVSFINSAKI---GGAIAIOSGSLSLAGESVLFQNNNSORT 321
      201 GCAIHITKNLTLSSCGETLFOGNAPTAAAGKGAIALADSCOTLISGDSGDIIFEGWT--1 258
      322 SDQGLV--RNAIYLEKALILSLAENG--DILFDPPI--VQESSKESPLPSLQASVTSPT 378
      259 GATGVSHSAIDGTSAKITLALRAAGHTTYEYDPIYTGISTVADAL-----NINSP- 311
      379 PATASPLVIQVTSANR---SVIFSSERLSE--EKKTPDNLTSOLOPTELKSGRIYKIDRA 433
      312 -----DTGNKKEVTCGLIVFSGEKLTEAKDEKNRSLKLVNVAFFNVYLVKQDV 362
      434 VLSAPSLSDPPALLINEACTSLKTSDD--LKLATSLPLSLDTEKSVTHA-----PNLS 488
      363 VLSANGFSQDANKRLINDLQTSLVANSTIELTNLEINIDSLRNGKKIKLSATAQKDIR 422
      489 IOKIFLNSGDEMPEYENVELLSKEONNIPLLTSLKEQSHLHPDGN-----LSSHGOGD 544
      423 IDRPVVALAISDESFTYQ--GFLNEDHSTGILELDAGKDVIASDSRSIDAVQSPYGTQK 481
      545 WPFESMKDSDGCHSLIANNTPKNVPPPEROSTLVANTLMTYSDMAVOSMINTIAHGCA 604
      482 WTIWNSITDOK--KATVSMAGQSFNPTAEQEAFLVPLNLWGSFIDVRSFQNFIELGTEGAP 539
      605 YLEGTWGSAVSNLFEYAHDSGKRPIDMMHRSILYFGISTHSLDDHSCFLAAQLLGS 664
      540 YERFVWAGISNVL--HRSGRNQKRFHVSGGAVVAGASTRMPGGDTLSLGFALFARDK 597
      665 DSPITSTETTSYIATVOAQLATSLMKISA-----QACYN 698
      598 DYEMNTLFAKTYAGSLRLQDASLYSVSLILBEGGLREILLPVYSKTLDCSFYQGLSYG 657
      699 ESIMELKTY-----RFSKGEFGSMHSAVSGCVASIPV--SNGSGLFSSSFISFKLQ 752
      658 HTDHRMKTESLPPRPILSLD--HTSMGCVYAGAGELGCRVAVENTSGGTFQDYTPFKVQ 716
      753 GFSGTQGFEESSGEIRSFASASSFRNLSLPIGTFEERKSQKTRTYFFLAGYIIDLKRDV 812
      717 AVYARODSFVELCAISRDFSDSHLYNLAIPLGILEKRF--AEQYHVAVMYSDDVCRSN 774
      813 ESGPNVLKNAVSADAPMANLDSRAVYFRLTNOALHRLQTL--NVSCVLRGQSHSYSLD 871
      775 PKCTTLLSNQSGWKTKGSLNARQGIYVQSGFSLGAAALDELGNFEMRSGSRSTYND 834
      872 LGTTYRF 878
      835 AGSKIKF 841

```


Oy 487 -LSIQIFLNSGDENFNEVAVELLSEKQNNIPILITLSEQSHLPD-----GNLSHF 539
 Db 572 ALTISCTGLVKNKSDCCDNHGHFNKDQVPILELKAITSNTVTTTDSLTGNGIQOSEY 631
 Oy 540 GYGDWTFESKDSDEGHSLIANWTPKNVPHPEROSTLVANTLWNTYSQMAVQSMINTI 599
 Db 632 GYOGTWEFTIDTFT--HTVGNKNKTKGYLPHPERLAPLIPKSLMANVYIDLRAVQA--SA 687
 Oy 600 AIG-----GAYLFTQMSAVNSLEYA-HDSSKRPIDNWHHBSLGLYFEGISTHSDHSEFL 654
 Db 688 AGGEDVPGKQLST---GTFNFHANITGDAR---SYRHMGGYLINTYTRITTPAALSL 741
 Oy 655 AAGQLLKSSDSFTSTSTETTSYIATVOAOLATSLMKIS-----AQACNESIHELKTK 707
 Db 742 GFGQLTKSKSDYLVGHGHSNVYFATVYSNITKSLFGSSRFSGTSRPTYSNSNEKVYS 801
 Oy 708 YRSFSKEGFSQMSVAVSGEVCASIPVNSGSL-FSSFSIFSKLQGESGTODGFEESG 766
 Db 802 YTKLPK-GRCSYNNCCNLGELFGLNLPYTLSSRLNLKQILPFVKAFAVAYATGCIQENRP 860
 Oy 767 EIRSFSSASRRNISLPGLITFEKKSKQTRITYYFLGAILQDLKRDVESGPVLLKNASVW 826
 Db 861 EGFIFGHILLNVAVPYGVAFRGKSNHRPDEYTIIVAAPVYRINPDCDTLLPLINGATW 920
 Oy 827 DAPMANLDSRAYMFRLTNORALHR-LQTLNVCGLVRCQSHSYSDLGQTYTRF 878
 Db 921 TSGNLTSTLLVQASSHTSVNDVLEIFGHGCGDIRTSNQYTLIDGSKLRF 973

 RESULT 12
 C81593
 polymorphic membrane protein G family CP0299 [imported] - Chlamydia pneumoniae (str
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #texel_change 11-May-2000
 C:Accession: C81593
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Guim, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A>Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; M01D:20150255
 A:Accession: C81593
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-995 <REA>
 A:Cross-references: GB:AE002191; GB:AE002161; NID:g7189216; PIDN:AAF8156.1; PID:g7189222
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP0299

	Query Match	18.0%	Score 816.5:	DB 2:	Length 995:
	Best Local Similarity	26.4%:	Pred. No. 1.5e-43:		
	Matches 252:	Conservative 150:	Mismatches 404:	Indels 139:	Gaps
Oy	NPNHVCNPFPEDCMESLFPCALAHAS---- <td>91</td> <td>:</td> <td>:</td> <td>:</td>	91	:	:	:
Dd	71 DPGRLCIFSDLLIANDLNIAISTSSCCSNRGAQLQILGKGVSFSLIRRSADGAAT	130	:	:	:
Oy	92 FK-----EKGDLSIONFRFLFTDCKSESPS----IIHQ-----NGOLSRRNGSMS	137	:	:	:
Dd	131 SSVITQNELCPLEBSFGSGSQMIFONCESLTLDSTPSASNVIPIHASAIYATPPTMLCTNNDSIL	190	:	:	:
Oy	138 FCRNHAECSGSAISADAFFSHOHNLYLTFAFENSSKGAGAIQ--AQFFSLSRNVSPISFAR	196	:	:	:
Dd	191 FOYNRSAGFAAILRGTSITTIETNTKKSLFPNGNGISINGCALTGSAINLIINSAPVIFT	250	:	:	:
Oy	197 NRADDNGAICCS--NLICGSNRNPVFLEFTGNSAINCAATICISLINISENG-----	245	:	:	:
Dd	251 NATGIYGAAIYLTCGSMITSCGNLSGVIVFNNSRSAGAIXANGWVPFSNNSDSLTFQONTA	310	:	:	:
Oy	246 -----SISLANOEFLFASNAKE	264	:	:	:
Dd	311 SPONSLEAPPRTPPPAVTPLLCYTGCALFCMPAPTTPPTTGTVALISCENSVEPLENIASE	370	:	:	:

Oy 265 KCGAYAAKHHVLRVNGVPSFPIINSAKIGCAIAI0SGSLSLAEGEVSLP0NNQ0RPSDQ 324
 Db 371 QGGALYGGKKISIDSNKSTIFLGNTAGKGGAIAIPESEGLSIA0G0ILEKKNLISITSGT 430
 Oy 325 GLVNNAIYLEKDAIILSLEARNQ-DILFDPPIVOE0SSKESPLPSLOASVSPPTATAS 383
 Db 431 P-TNHSIHFCKDAKFATLGATQGYLVFYDPTISDDLSAAS-----AAA 473
 Oy 384 PLVYQTSANR-----SYIFSSERL-SEEEKTP0NLTSOLOQPIELKSGRLVKDRAVLS 436
 Db 474 TVVNVNPKASADGAYSGYIVFSGEFLTATEAATPANANSTLNOKELEGGFALNCGATLNL 533
 Oy 437 APSLSODQALLIIEAGISLTKSSD-----LKLATISLPLSLDPEKSVTIAHP- 486
 Db 534 VHNFTQDCKSVYINDAGITLTLTTNGANNTDGAILNKLVIYLDSDLGKAAVNV0STNG 593
 Oy 487 -LSIQIKFLNSGDENFYENVELLSKEONNPLITLSKE0SHLHPD-----GNLSHF 539
 Db 594 ALTISGLGLVKNQSDCCDNHGMFNKLOQVPILELKATSMVTYTTDPSLGTNGYQ0SPY 653
 Oy 540 GYCGDWTFSMKDSDEGHSILANMTPKNVPIPEKOSTLVANTLNTYSDMOAV0SMINTI 599
 Db 654 GYCGWETITDPTT-HVVTGNMKKTYGLPPIPERLAPILPMSLVANVYIDIAVVSQA-SA 709
 Oy 600 AHG-----CAYLEFGMGSAVSNLFLYA-HDSGKPI0NMHNRSLGLFGISTSLDHSRCL 654
 Db 710 ADGEDVPEK0LST---GITNFFHANNITGDAR---STRHMGGGLINTYTRITPDALSL 763
 Oy 655 AAGQLLGKSSDSFITSPETTSYIATVQALATSLMKIS-----AQACVNESIHELKTK 707
 Db 764 GEGQLFTMSKQYLVGHGHSNVYFATVYSNLTKSLFSGSRFPSSGTSRYTYSRSMEKYKTS 823
 Oy 708 YRSTSKESFGSMHSAVNSGEVCASTPIYVSN0SGL-FSSFSIFSKL0GFSGTQDPESSG 766
 Db 824 YTKLPR-GRCSWNNCWLGELEGNLPIILSRILNLKOIIPFVKAEEVAYATHGIGQENTP 882
 Oy 767 EIRSPSSAPNSHISLIGITFEKKSQKTRTYVYFAGVYI0DLKRVESGPAVLLKNAVSM 826
 Db 883 EGRIFGHGHLNVAVPVYVRECKNSHNRPDYTIIVAAAPDYVYKHNPCDPTLPLNGATW 942
 Oy 827 DAPMANLDSRAVYFRLTNOALHR-LQTLNVSCVLRG0SHSYSIDIGTYYRF 878
 Db 943 TSGINLRLRSTLIVQASSHTSYNDVLEIFGHGCGDIRTSQYTLIDISKLRF 995

RESULT 13
 C72078
 polymorphic outer membrane protein g family - Chlamydothila pneumoniae (strain CWL0299
 C.Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
 C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
 C.Accession: C72078
 R.Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
 Nature Genet. 21, 385-389, 1999
 A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A.Reference number: A72000; MUID:99206606
 A.Accession: C72078
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-936 <AR>
 A.Cross-references: GB:AE001627; GB:AE001363; NID:q4376721; PIDN:AAD18589.1; PID:q4376721
 A.Experimental source: strain CWL0299
 C.Genetics:
 A:Gene: pmp_7

Query Match	773	DB 2	Length	936
Best Local Similarity	28.0%	Pred.	No. 7, 8e-41	
Matches	236	Conservative	143	Mismatches 376; Indels 86; Gaps 22;

QY	97	D	S	I	O	N	F	R	L	S	T	D	S	S	K	E	S	P	S	--	I	I	H	O	K	N	Q	L	S	L	R	N	N	G	S	M	S	P	F	R	N	A	E	G	G	A	I	S	D	A	154
Db	119	N	L	F	N	D	F	S	R	L	S	T	S	P	L	S	P	T	G	O	A	L	K	S	V	O	N	L	S	T	G	S	I	I	L	P	O	N	S	S	D	G	G	J	I	N	K	N	178		

Db 811 NFKSGAEARAFDDGDLVNCSTIPVIGIRLEKISEDEKNNFELSLAYIDVYRKNNRSTSL 870
 Qy 820 LKNAVSDAPMANLDSRAVYFRLNQRALH-RIOTLLNVSCVLRCGSHSYSLDGTYYRF 878
 Db 871 MVSASWTSLSCKNLARQAFALASAGSHLTLSPHVELSGEAAVELGSAHYIVNDCGLRYSF 930

RESULT 5

D72078

polymorphic outer membrane protein g family - Chlamydia pneumoniae (strain CWL029)

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C:Accession: D72078

R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: D72078

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-930 <ARN>

A:Cross-references: GB:AE001627; GB:AE001363; NID:94376721; PIDN:AAU18590.1; PID:9437672

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: pmp_8

Query Match 20.28; Score 915.5; DB 2; Length 930;

Best Local Similarity 30.8%; Pred. No. 7.6e-50;

Matches 255; Conservative 150; Mismatches 344; Indels 87; Caps 26;

Qy 98 LSTIONFFRLS-----TDCSSKSSPSIIHQKQGLSRNNSMFCR--NIAEGSGA 149

Db 119 LTFGCFNLSPFIAPGTTVASGKSTLS--SAGALNLTQNCITLFSQVNSNANNGCA 174

Qy 150 ISADAFSLQNHVLTFAEENSSKNGCAI-QAOTFSLSRNVPISPARNRADLNGALIC 208

Db 175 ITTKTSLISGNTSITFTSSAKRLGAIYSSAASISGNGOLVFMNKKETGGAL-- 232

Qy 209 SNLICSGNV--NPULFTGNSATN---GCAICISDINTSEKSLSLACQETLFAFNS 261

Db 233 -GFEASSITQNSSLFTSGNATDAKGAHYC---EKTEFTPLTISGKSLTFAENS 288

Qy 262 AKERGAIAKHAHVLRNGPVSTINN---SAKIGAIAIQSGSLSILAGECVLEFON 316

Db 289 SVTGGALICAHGLDLSAAGPLTFSSNMGNTAAGKALAIADGSLSLANOCDDIFELG 348

Qy 317 NS-QRTSOGVRAIYLEKAILLSLEARNG-DILFPDPIVQSSSKESPSPSLQASV 374

Db 349 NTLTSTSPSTRNAITLGSAAKITNLRAGGOSITFFDPIASNTGASDVL----- 400

Qy 375 TSPPIATASPLVITQTSANRSYIPSSERLS--EEKTPDNLTSOLOQPIELKSGRLVLDRA 433

Db 401 TINDPDSNPL---DYSGITVFSGEKLSADEAKADNFTSLKOPALASGTLALAGCNV 456

Qy 434 VLSAPLSQDPOALLIMEAGTSLKTS--LKLATLSIPPLHSIDREKSVTHAP-----N 486

Db 457 ELVDNGCFOTGSLTLQKPTKLDATKALSLTKLVVDLSLEGNKSVSITAGANKTIT 516

Qy 487 LSIQKIFLNSGDNENFENVELLSKEQNNIPLLLSKQSHLHPDGLLS-----HF 539

Db 517 LTPSLVLEDDSSG--NFESHTINQAFQPLVFAATFASDIYI-DALLTSVPVOTPEPHY 573

Qy 540 GYGQDMTFSMKDSDEGSHLIANNTPKVNVPHERQSPFLVANTLNNVSDMAVOSMINTI 599

Db 574 GYGQHMEATMADITSTAKSGTWTVTGYNPRERASVPSLSASFDDITDQIMISQ 633

Qy 600 AHGAIVLEFTWGSANLFLYHDSGKPIDMHHRSGLYFGISTHSLDHSFCLAQOL 659

Db 634 ANSIVYQOGLASGTANFFH-KDKSGTN-QAFRHSYCYIVGSAEDSESENIFFAVCOL 691

Qy 660 LKGSDSITSTETTSYATATQAO-----LATSLEK-----ISAQACYNES 700

Db 660 LKGSDSITSTETTSYATATQAO-----LATSLEK-----ISAQACYNES 700

Db 692 FGKDKDLFIVENTSHNLTASLYLQHRAFGLPMPSPGSTITDMKIDPLILNQLSTYT 751
 Qy 701 IHELTTRYRSPFSGEFGSMHSAVSGEVCASIP1-VNSGQLFSSPSIFSKLOGFSCTQD 759
 Db 752 KNDMDTRTYTSV-PEAQGSWMNNSCALFELGSLALYLKKAEPFGCYRPPLKFAVYSRQ 810
 Qy 760 GFEESGSEIRSFSSASRRNLSLPITGITEKKSQKTRYYFLGAYIDLKRVDSEGPVL 819
 Db 811 NFKSGAEARAFDDGDLVNCSTIPVIGIRLEKISEDEKNNFELSLAYIDVYRKNNRSTSL 870
 Qy 820 LKNAVSDAPMANLDSRAVYFRLNQRALH-RIOTLLNVSCVLRCGSHSYSLDGTYYRF 878
 Db 871 MVSASWTSLSCKNLARQAFALASAGSHLTLSPHVELSGEAAVELGSAHYIVNDCGLRYSF 930

RESULT 6

C81591

polymorphic membrane protein g family CP0309 [imported] - Chlamydia pneumoniae (s

C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: C81591

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

Nucleic Acids Res. 28, 1397-1406, 2000

A:Reference number: AB1500; MUID:20150255

A:Accession: C81591

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1276 <REA>

A:Cross-references: GB:AE002193; GB:AE002161; NID:97189224; PIDN:AAF36166.1; PID:9718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0309

Query Match 19.78; Score 894.5; DB 2; Length 1276;

Best Local Similarity 30.6%; Pred. No. 2.7e-48;

Matches 279; Conservative 144; Mismatches 364; Indels 125; Caps 29;

Qy 68 LTVLGNSTCMVSKL-----HITDPKLEKKEKGLDLSIONRFLSPDSCSKSSPSIIH 122

Db 389 LYSKGNSSLTGNTNLLFSGNKATGPSNNSANQEG-----CGAILSTLESAS-----VS 437

Qy 123 QKNQGLSRNNSMFCRNHAEGSGAISADAFSLQNHVLTFAEENSSKNGCAIQAOT 182

Db 438 TKKG-LWIEDENENVSLSCNATVSGAIVATKALHGNNTLIT-PDGNTEFAGAIYET 495

Qy 183 --FSLSRNVPISPARNRADLNGALICCSNLICGNNVPLEFTNSATN----- 229

Db 496 EDFTLTGSYGTIVFTSTAKTAGALHTKGNTPFKN-KALVFSGNSAATATTTTDOEGC 554

Qy 230 GGAICCC-ISPDLTSEKSLSLACQETLFAFNSAKE----- 264

Db 555 GGALICNISESDITK-SLTITENESLFTNNTKKRGGGIYAKCVISGSESTINFGNT 613

Qy 265 --KGAIAKHAHVLRNGPVSTINSAKTOGAIAIQSGSLSILAGECVLEFONNSORT 321

Db 614 AETSQGAIVSKNLSITFANGPVSTFNNSGCGKAIYIDSGELSLAIDGITFGCN--RA 671

Qy 322 SDQGLVRAIYLEKDALLSLEARNG-DILFPDPIVQSSS-----KESPLPSLSQSVT 375

Db 672 TEGTSTPNSHLGAARITKLAAPGHTIYFDITMEAPASGCTIEELVNPVKAIVP 731

Qy 376 SPT-----PATASPLVITQTSAN--RSVIFSSERL-SEEEKTPDNLTSOLOQPIELKSGRL 427

Db 732 PPQKNGPIASVPVPAFANPNNGITVIFSSGKLPSQDASIPANTTTILNOKIKLAGNV 791

Qy 428 VLKRAVLSAPLSQDPOALLIMEAGTSLKTS-----DLKATLSIPPLHSIDREKSVT 481

Db 792 VLKGCATLQVYSFTQQPDSTVFMDAGTLETETTTNNNDGSLDLNLSVNDALDGKMIT 851

Qy 482 I----HAPNLSIQKIFLNSGDNENFENVELLSKEQNNIPLLTSLKQSHLHPDGN--- 534

Db 482 I----HAPNLSIQKIFLNSGDNENFENVELLSKEQNNIPLLTSLKQSHLHPDGN--- 534

OY	421	ELKSRVLKRAVLSAPSLSDPOALLIMEAGTSLKTSDDLKATLTIPLHSLDTEKSV	480
Db	421	ELKSRVLKRAVLSAPSLSDPOALLIMEAGTSLKTSDDLKATLTIPLHSLDTEKSV	480
OY	481	TIHAPNLSIOKIFILSNSDENFENVELLSKEONNIFPLTTSKEOSHLPDGLNLSHFG	540
Db	481	TIHAPNLSIOKIFILSNSDENFENVELLSKEONNIFPLTTSKEOSHLPDGLNLSHFG	540
OY	541	YOGDWTFSMKSDDEGHSLLANWTPKNVYPHPEROSTLVANTLWNTYSDMQAVQSMINTIA	600
Db	541	YOGDWTFSMKSDDEGHSLLANWTPKNVYPHPEROSTLVANTLWNTYSDMQAVQSMINTIA	600
OY	601	HGGAFLPCTHMSAVSNLFLYAHDSGSKPIDNMHMSLGLFGISHSILDDHSFCLAAQOLL	660
Db	601	HGGAFLPCTHMSAVSNLFLYAHDSGSKPIDNMHMSLGLFGISHSILDDHSFCLAAQOLL	660
OY	661	GKSSDSFTTSTETTSYIAVTVOAOLATSLMKISAPACYNESIHELKTYRSPSKEGFSGSWH	720
Db	661	GKSSDSFTTSTETTSYIAVTVOAOLATPLMKISAPACYNESIHELKTYRSPSKEGFSGSWH	720
OY	721	SVANSCEVCASIPVYNSNGCLFSSFSJFSKIQGFSGTODGCEBSGELTRSFSAASSFRNIS	780
Db	721	SVANSCEVCASIPVYNSNGCLFSSFSJFSKIQGFSGTODGCEBSGELTRSFSAASSFRNIS	780
OY	781	LPIGITFEKSKOKRTTYYYFLGAVIQDLKRDVEGSPVLLKNAYSMDBPAMNLDSRAFM	840
Db	781	LPMGITFEKSKOKRTNYYYFLGAVIQDLKRDVEGSPVLLKNAYSMDBPAMNLDSRAFM	840
OY	841	RLTNORALHRLQTLNLNASCVLRCQSHSYSLDLGTTYR	878
Db	841	RLTNORALHRLQTLNLNASCVLRCQSHSYSLDLGTTYR	878

RESULT 2

F81721

polymorphic membrane protein G family TC0267 [imported] - Chlamydia muridarum (strain Nijmegen)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: F81721

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldoberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L. Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255

A:Accession: F81721

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-867 <TEN>

A:Cross-references: GB:AE002294; GB:AE002160; NID:g7190305; PIDN:AAF39136.1; PID:g7190305

A:Experimental source: Strain Nig9 (MoPn)

C:Genetics:

A:Gene: TC0267

	Query Match	78.28;	Score 3543.5;	DR 2;	Length 867;
	Best Local Similarity	77.88;	Pred. No. 5.7e-216;		
	Matches 684;	Conservative 76;	Mismatches 106;	Indels 13;	Gaps 2;
Qy	1 MRPDHNFCCCAALISSTAVLFCODPLGETALLTKNPNHVCFFEDCTMESLFPALCA 60				
Db	1 MRPDHVNLCCLCATILSTPAILFGODALDKSLITKNNNSIVCFLEEDCTMENSPLALS 60				
Qy	61 HASDDPLVLCNLSYCMFVSKLHITDPKEALFKKEGDLSTIONFRILSTPTDCSSK-ESSPS 119				
Db	61 HARODDPLVLIIGTNTMHWVSNLHPSTNEERFLKEKGDLSTIDFRLSTPTDCSSSEUDPS 120				
Qy	120 ITHKNGQLSTLNNNGSMFCRHHAGSGAISADAFSLSHYHNLFTFAFEENSSKGGGAIQ 179				
Db	121 ILYHKNGQLFLRNNNMMSFYRRHSGSGALSTDLALFLQHMYLFTFNEENSAAKNGGAIQ 180				
Qy	180 AQTSLSRNVSPISFARNRADIINGCAICCSNLIICSGVNPVLPFTGNSATNGAICCIIDL 239				
Db	181 AQTSLSRNVSSLSFARNRADIINGCAICCSNLIICSGVNPVLPFTGNSATNGAICCIIDL 240				

[illegible]

RESULT 3
B72077
polymorphic membrane protein G family CP0306 [imported] - Chlamydothrix pneumoniae (s
C.Species: Chlamydothrix pneumoniae, Chlamydia pneumoniae
C.Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C.Accession: B72077, B81592
R.Katman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,
Nature genet. 21, 385-389, 1999
A.Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A.Reference number: A72000; MUID:99206606
A.Accession: B72077
A.Molecule type: DNA
A.Residues: 1-928 <ARN>
A.Cross-references: GB:AE001628; GB:AE001363; NID:94376730; PIDN:PAD18591.1; PID:94377
A.Experimentals: GB:AE001628; strain CML029
R.Reed, T.D.; Brumham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey
C.; Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1997-1406, 2000
A.Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR399
A.Reference number: AB1500; MUID:20150255
A.Accession: B81592
A.Molecule type: DNA
A.Residues: 1-928 <REA>
A.Cross-references: GB:AE002192; GB:AE002161; NID:97189226; PIDN:AAF38163.1; PID:97188
A.Experimental source: strain AR39, HL cells
C.Genetics:
A:Gene: pmp_-9; CP0306